REMARKS

Claim Rejections under 35 U.S.C. § 101
Utility:

The Examiner has maintained the rejection of claims 25-36 under 35 U.S.C. § 101, alleging that the claimed invention is not supported by either a substantial asserted utility or a well established utility. Specifically, the Examiner does not find persuasive the argument that the claims are supported by the utility of creating degenerative oligonucleotide probes for isolation of genomic and cDNA sequences that are amplified in lung and colon tumors. Additionally, the Examiner does not find Applicants arguments based on the utility asserted in the specification at pages 119 and 137, and further supported by the Goddard and Ashkenazi declarations, to be persuasive.

More specifically, the Examiner rejects Applicants' utility arguments because according to the Examiner, in the "absence (of) evidence that the polypeptide is expressed at an elevated level, one (of ordinary skill in the art) would conclude that the claimed invention is not supported by either a substantial asserted utility or a well-established utility." Office Action mailed February 5, 2004. In particular, the Examiner maintains that the amplification of SEQ ID NO:68 does not provide a readily apparent use for the polypeptide because there is no information regarding the level of expression, activity, or role in cancer of SEQ ID NO:69.

The Examiner relies on Pennica et al., as support for finding the present invention lacks utility in the absence of evidence of overexpression of the claimed polypeptide. Specifically, the Examiner alleges that Pennica et al. provides an example of copy number being amplified but RNA expression actually being reduced.

Applicants respectfully disagree with the Examiner's assertion that absent evidence of overexpression of the claimed polypeptide the present invention is not supported by a utility. First, in rejecting both Applicants' assertions of utility, and the Goddard and Ashkenazi declarations, the Examiner has set the standard for satisfying the utility requirement too high. Under the proper utility standard, Applicants have demonstrated

2

that the present invention is supported by a specific, substantial, and credible utility. Specifically, Applicants herein cite several art references which indicate that one of ordinary skill in the art would not have reasonably questioned the utility asserted at pages 119 and 137 of the specification. Second, the reference relied on by the Examiner, Pennica et al., does not outweigh the evidence Applicants submit as support demonstrating that those of skill in the art would reasonably expect the protein expression levels of the claimed polypeptides to correlate to the amplified levels of DNA. Third, consistent with the Utility Guidelines, Applicants have demonstrated that the present invention is supported by a specific, substantial, and credible utility.

A.   The Examiner Sets the Utility Bar Too High

As Applicants have previously argued, at pages 119 and 137 of the specification, Applicants assert a specific, substantial, and credible utility for the claimed invention:

> Amplification is associated with overexpression of the gene product, indicating that the polypeptides are useful targets for therapeutic intervention in certain cancers such as colon, lung, breast and other cancers. Therapeutic agents may take the form of antagonists of PRO327, PRO344, PRO347, PRO357 aor (sic) PRO715 polypeptide, for example, murine-human chimeric, humanized or human antibodies against a PRO327, PRO344, PRO347, PRO357, or PRO715 polypeptide. These amplifications are useful as diagnostic markers for the presence of a specific type of tumor.

(p.119)

> The polypeptides encoded by the DNAs tested have utility as diagnostic markers for determining the presence of tumor cells in lung and/or colon tissue samples.

(p.137)

An Applicant's assertion of utility creates a presumption of utility sufficient to satisfy the utility requirement of 35 U.S.C. § 101, "unless there is a reason for one skilled in the art to question the objective truth of the statement of utility or its scope." *In re Langer*, 183 USPQ 288, 297 (CCPA 1974). *See also In re Jolles*, 206 USPQ 885 (CCPA 1980); *In re Irons*, 144 USPQ 351 (9165); *In re Sichert*, 196 USPQ 209, 212-213 (CCPA 1977).

Compliance with 35 U.S.C. § 101 is a question of fact. *Raytheon v. Roper*, 724 F.2d 951, 956, 220 USPQ 592, 596 (Fed. Cir. 1983) *cert. denied*, 469 U.S. 835 (1984). The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the totality of the evidence under consideration. *In re Oetiker*, 977 F.2d 1443, 1445, 24 USPQ2d 1443, 1444 (Fed. Cir. 1992). Thus, to overcome the presumption of truth that an assertion of utility by the Applicant enjoys, the Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility.

Further, statistical certainty regarding Applicants assertion of utility is not required to satisfy 35 U.S.C. § 101. *Nelson v. Bowler*, 626 F.2d 853, 856-857, 205 USPQ 881, 883-884 (CCPA 1980). Where an Applicant has specifically asserted that an invention has a particular utility, that assertion cannot simply be dismissed as "wrong" even where there may be some reason to question the assertion. MPEP § 2107.02. Significantly, a 35 U.S.C. § 101 rejection should only be sustained where the asserted utility violates a scientific principle or is *wholly* inconsistent with contemporary knowledge in the art. *In re Gazave*, 379 F.2d 973, 978, 154 U.S.P.Q. 92, 96 (CCPA 1967) (emphasis added).

Consideration of the totality of the evidence discussed below clearly demonstrates that the proposition that there will be correlation between protein and transcript levels does not violate scientific principles nor is it wholly inconsistent with knowledge in the art. Thus, the maintained rejection of the present claims for lack of utility is improper and should be withdrawn.

1. **It is a general scientific principle that DNA is transcribed into RNA which is translated into protein.**

According to Genes V, a central dogma of molecular biology is that genes are perpetuated as nucleic acid sequences, but function by being expressed in the form of proteins. Thus, genetic information is perpetuated by replication where a double-stranded nucleic acid is duplicated to give identical copies. These copies are then expressed by a two-stage process. First, transcription generates a single-stranded

RNA identical in sequence with one of the strands of the duplex DNA. This RNA strand is then translated such that the nucleotide sequence of the RNA is converted into the sequence of amino acids comprising a protein. *See* Lewin, Bejamin. *Genes V.* 1994. Oxford University Press, NY, NY. p. 163. (Appendix A).

Thus, those of skill in the art generally accept that gene expression levels correlate to protein expression levels absent specific events such as translation regulation, post-translation processing, protein degradation, protein isolating errors, etc. *See* Orntoft *et al.,* "Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas." 2002. *Molecular & Cellular Proteomics 1.1,* 37-45. (Appendix B). Therefore, Applicants' assertion that the claimed polypeptides are supported by a diagnostic utility because they are encoded by nucleic acids that are amplified in lung and colon tumors does not violate scientific principles.

2.   Utility of the Claimed Polypeptides is Not Wholly Inconsistent with Knowledge in the Art

Pollack, Orntoft, Hyman, Bermont, Varis, and Hu demonstrate that the utility of the claimed polypeptides is not wholly inconsistent with the knowledge in the art. These references further support Applicants' argument that one of ordinary skill in the art would reasonably conclude that the present invention is supported by a specific, substantial, and credible utility.

For example, Pollack et al. profiled DNA copy number alterations across 6,691 mapped human genes in 44 breast tumors and 10 breast cancer cell lines and reported that microarray measurements of mRNA levels revealed remarkable degrees to which variation in gene copy number contributes to variation in gene expression in tumor cells. *See* Pollack et al., "Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors." 2002. *PNAS,* 99(20):12963-12968. (Appendix C). Pollack et al further report that their findings that DNA copy number plays a role in gene expression levels are generalizable. Thus significantly, "[t]hese findings provide evidence that widespread DNA copy number

alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer."

In particular, Pollack et al. report a parallel analysis of DNA copy number and mRNA levels. Pollack et al. found that "[t]he overall patterns of gene amplification and elevated gene expression are *quite concordant, i.e.,* a significant fraction of highly amplified genes appear to be correspondingly highly expressed." (emphasis added). Specifically, of 117 high-level DNA amplifications 62% were associated with at least moderately elevated mRNA levels and 42% were found associated with comparably highly elevated mRNA levels.

Orntoft et al report similar findings in "Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas." 2002. *Molecular & Cellular Proteomics 1.1,* 37-45. (Appendix B). Initially, Orntoft et al. note that "[h]igh throughput array studies of the breast cancer cell line BT474 ha(ve) suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas ( ), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, *cyclin d1, ems1,* and N-myc."

Specifically, Orntoft et al. used 2D-PAGE analysis on four breast tumor tissue samples to determine correlation between genomic and protein expression levels of 40 well resolved, known proteins. Orntoft reported that "[i[n general there was a *highly significant correlation* (p<0.005) between mRNA and protein alterations ( ). Only one gene showed disagreement between transcript alteration and protein alteration." (emphasis added). Additionally, Orntoft et al. report that "11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level ( )." The regions examined by Orntoft include genes encoding proteins that are often found altered in bladder cancer.

Orntoft et al. note that their study reports a *striking correspondence* between DNA copy number, mRNA expression and protein expression. Orntoft et al., further note that any observed discrepancies in correlation may be attributed to translation regulation, post-

6

translation processing, protein degradation or some combination of these. *See also* Hyman *et al.,* "Impact of DNA amplification on gene expression patterns in breast cancer." 2002. *Cancer Research*, 62:62-40-6245. (Appendix D).

Varis and Bermont are yet further examples that utility of the present invention based on a correlation between gene amplification and protein overexpression is not wholly inconsistent with knowledge in the art. Varis et al., carried out a comprehensive analysis of gene copy number and expression levels of 636 chromosome 17-specific genes in gastric cancer. *See* Varis et al., "Targets of gene amplification and overexpression at 17q in gastric cancer." *Cancer Res.* 2002. 1;62(9):2625-9. (Appendix E). Specifically, Varis et al report that analysis of DNA copy number changes by comparative genomic hybridization on a cDNA microarray revealed increased copy numbers of 11 genes, 8 of which were found to be overexpressed in the expression analysis. Thus, Varis et al., teach there is a 72% correlation between increased DNA copy number and gene expression level.

Bermont teaches that overexpression of p185 is usually associated with c-erbB-2 amplification. Specifically, Bermont reports that 100% of the overexpressed p185 protein in 106 breast cancer samples studied also displayed c-erbB-2 amplification. *See* Bermont et al., "Relevance of p185 HER-2/neu oncoprotein quantification in human primary breast carcinoma." *Breast Cancer Res Treat.* 2000 63(2):163-9. (Appendix F). *See also* Hu et al., "Profiling of differentially expressed cancer-related genes in esophageal squamous cell carcinoma (ESCC) using human cancer cDNA arrays: overexpression of oncogene MET correlates with tumor differentiation in ESCC." *Clin Cancer Res.* 2001 7(11):3519-25 (the results of cDNA arrays showed that 13 cancer-related genes were upregulated > or = 2 fold and immunostaining results of the expression of the MET gene showed MET overexpression at the protein level, validating the cDNA arrays findings). (Appendix G).

Thus, although there may not always be a 100% correlation between gene amplification and protein overexpression, the above discussed references evidence that the utility of the present invention is not wholly inconsistent with the knowledge in the art, and

therefore, also evidence that one of ordinary skill in the art would believe the claimed invention to be supported by a specific, substantial, and credible utility.

B. Pennica et al. Does Not Outweigh the Teachings of the Specification and the References Cited by Applicants

The Examiner argues that Pennica et al. provides an example where DNA copy number is amplified but mRNA expression is actually reduced. Applicants respectfully disagree.

Pennica et al. recognize that "amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance." For this reason, Pennica et al. analyzed relative gene amplification and RNA expression of *WISPs-1, 2,* and *3* in cell lines, colorectal tumors, and normal mucosa using quantitative PCR.

Initially, Pennica et al. noted that *WISPs-1* and *2* had copy numbers that were significantly higher than one, indicating gene amplification. Pennica et al. further noted that the copy number for *WISP-3* was "indistinguishable" from one (p=1.666), indicating no or minimal gene amplification.

Next, Pennica et al. examined the levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa using quantitative PCR. Pennica et al. found that *WISP-1* RNA levels displayed *good correlation* to gene amplification of *WISP-1*. Specifically, Pennica et al. found that RNA levels of *WISP-1* in tumor tissue were significantly increased in 84% (16/19) of the human colon tumors examined when compared with normal adjacent mucosa. *See* page 14721, Figure 7.

However, Pennica et al. also found that *WISP-3* RNA levels did not significantly correlate with *WISP-3* gene amplification. In particular, although *WISP-3* did not display significant gene amplification levels, RNA levels in tumor tissue were overexpressed in 63% (12/19) of the human colon tumors examined when compared with normal adjacent mucosa.

Further, as the Examiner notes, Pennica et al. also report that *WISP-2* gene amplification levels are inversely correlated with RNA expression levels. That is,

although *WISP-2* was significantly amplified, RNA levels of *WISP-2* in tumor tissues were significantly lower than RNA levels of *WISP-2* in normal adjacent mucosa. Specifically, 79% (15/19) of the samples examined displayed this expression pattern.

The Examiner relies on this last result as support for the proposition that one of ordinary skill in the art would not expect gene amplification levels to correlate with protein overexpression absent explicit evidence of protein overexpression. Applicants respectfully disagree.

First, *WISP-1* gene amplification and RNA expression levels showed a significant positive correlation. Second, although *WISP-3* was not significantly amplified, it was amplified (*P*=1.666) and significantly overexpressed. Third, although *WISP-2* gene amplification and RNA expression levels seemed to be inversely related, Pennica et al. state that this result might be inaccurate. Specifically, Pennica et al. suggest that "[b]ecause the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for *WISP-2* may be caused by another gene in this amplicon." *See* 14722. Thus, because the RNA expression pattern of *WISP-2* cannot be accurately attributed to gene amplification of *WISP-2,* this result should be disregarded. Therefore, particularly in light of the references discussed above, one of ordinary skill in the art may conclude that Pennica et al. supports a utility for the present invention because Pennica et al. teaches that gene amplification of *WISP-1* strongly correlates (84%) with RNA overexpression.

C. The Claimed Invention is Supported by a Utility that is Specific, Substantial, and Credible

Finally, use of the polypeptide sequence of PRO357 as a diagnostic marker is a specific, substantial and credible utility.

"Specific utility" is defined as:

> [a] utility that is *specific* to the subject matter claimed. This contrasts with a *general* utility that would be applicable to the broad class of the invention. For example, a polynucleotide whose use is disclosed simply as a 'gene probe' or 'chromosome marker' would not be considered to be *specific* in the absence of a

9

disclosure of a specific DNA target. Similarly, a general statement of diagnostic utility, such as diagnosing an unspecified disease, would ordinarily be insufficient absent a disclosure of what condition can be diagnosed.

Revised Interim Utility Guidelines Training Materials, pgs. 5-6 (http://www.uspto.gov/ web/offices/pac/utility/utilityguide.pdf). The presently claimed polypeptides are asserted to be useful as targets for therapeutic intervention in lung or colon cancer or as diagnostic markers, indicating the presence of tumor cells in lung or colon tissue samples. These utilities are specific to the claimed polypeptides, which are encoded by nucleic acids that are amplified in lung or colon tumors and therefore, the claimed polypeptides are supported by a specific utility.

"Substantial utility" is defined as:

> a utility that defines a 'real world' use. Utilities that require or constitute carrying out further research to identify or reasonably confirm a "real world" context of use are not substantial utilities. For example, both a therapeutic method of treating a known or newly discovered disease and an assay method for identifying compounds that themselves have a "substantial utility" define a "real world" context of use. An assay that measures the presence of a material which has a stated correlation to a predisposition to the onset of a particular disease condition would also define a "real world" context of use in identifying potential candidates for preventive measure or further monitoring.

Revised Interim Utility Guidelines Training Materials, pg. 6http://www.uspto.gov/web/ offices/pac/utility/utilityguide.pdf). The presently claimed polypeptides are also supported by a substantial utility because the utilities discussed above, *i.e.* therapeutic targets and diagnostic markers, are real world uses. For example, similar to the statement found in the above quote from the Guidelines, the present specification discloses an assay that measures gene amplification in cancerous cells. The articles discussed earlier correlate that gene amplification in cancerous cells with polypeptide overexpression in cancerous cells. Therefore, the claimed polypeptides are supported by a substantial utility.

"Credible utility" is defined as:

> Where an applicant has specifically asserted that an invention has a particular utility, that assertion cannot simply be dismissed by Office personnel as being

10

'wrong'. Rather, Office personnel must determine if the assertion of utility is credible (*i.e.*, whether the assertion of utility is believable to a person of ordinary skill in the art based on the totality of evidence and reasoning provided). An assertion is credible unless (A) the logic underlying the assertion is seriously flawed, or (B) the facts upon which the assertion is based are inconsistent with the logic underlying the assertion. Credibility as used in this context refers to the reliability of the statement based on the logic and facts that are offered by the applicant to support the assertion of utility. A *credible* utility is assessed from the standpoint of whether a person of ordinary skill in the art would accept that the recited or disclosed invention is currently available for such use. For example, no perpetual motion machines would be considered to be currently available. However, nucleic acids could be used as probes, chromosome markers, or forensic or diagnostic markers. Therefore the credibility of such an assertion would not be questioned, although such a use might fail the *specific* and *substantial* tests.

Revised Interim Utility Guidelines Training Materials, pg. 5 (http://www.uspto.gov/web/ offices/pac/utility/utilityguide.pdf). The present invention is supported by a credible utility. As discussed earlier at pages 5-8, the references cited by Applicants demonstrate that the logic underlying Applicants assertion of utility is not seriously flawed, nor are the facts upon which utility is asserted inconsistent with the logic underlying the assertion of utility. Therefore, utilizing the claimed polypeptides as therapeutic targets or diagnostics markers in lung or colon cancer is a credible utility.

For all the above reasons, Applicants have demonstrated currently pending claims 25-36 are supported by an asserted substantial, specific, and well-established utility and therefore, respectfully request that the Examiner withdraw the rejection of claims 25-36 for lack of utility.

35 U.S.C. § 112 ¶ 1, Enablement-Utility

The Examiner has rejected claims 25-36 under 35 U.S.C. § 112 ¶1, alleging that because the claimed invention is not supported by either a specific asserted utility or a well established utility, one skilled in the art would not know how to use the claimed invention. As discussed in the remarks above, addressing the rejection under 35 U.S.C. § 101 for lack of utility, Applicants respectfully submit that the claimed polypeptide is supported by a specific, substantial, and credible utility. Thus, Applicants respectfully request the Examiner reconsider and withdraw the rejection of claims 25-36 under 35

11

U.S.C. § 112 ¶1 for their alleged inadequate disclosure on how to use the claimed invention.

Claim rejections under 35 U.S.C. § 112, first paragraph
Written Description:

The Examiner has maintained his rejection of Claims 25-26 and 33-34 under 35 U.S.C. § 112, first paragraph, contending that they contain subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention.   Applicants respectfully disagree.

First, the Examiner argues that "it seems clear that applicant is not in possession of the claimed invention because the response admits that one 'might also be isolated' which indicates they were not isolated at the time of the claimed invention." Office Action mailed June 8, 2004, pages 4-5. Applicants respectfully disagree that this is a proper ground for rejection and disagree that Applicants' statement can be used as indicia of lack of written description. Specifically, "[a]pplication of the written description requirement ( ) is not subsumed by the 'possession' inquiry." *Enzo Biochem, Inc. v. Gen-Probe, Inc.,* 323 F.3d 956, 961, 63 USPQ2d 1609, 1617 (Fed. Cir. 2002).

In fact, according to MPEP § 2163.02, "possession" of an invention may be shown in many ways, only one of which requires actual physical possession or reduction to practice of the claimed invention. For example, "possession" may be shown by "showing that the invention was 'ready for patenting' such as by the disclosure of drawings or structural chemical formulas that show the invention was complete, or by describing distinguishing identifying characteristics sufficient to show that the applicant was in possession of the claimed invention. *See, e.g., Pfaff v. Wells Electronics, Inc.,* 525 US 55, 68, 119 S. Ct. 304, 312, 48 USPQ2d 1641, 1647 (1998); *Regents of the University of Calif. v. Eli Lilly,* 119 F.3d 1559, 1568, 43 USPQ2d 1398, 1406 (Fed. Cir. 1997); *Amgen, Inc. v. Chugai Pharmaceutical,* 927 F.2d 1200, 1206, 18 USPQ2d 1016,

12

1021 (Fed. Cir. 1991) (one must define a compound by 'whatever characteristics sufficiently distinguish it')."

According to MPEP § 2163 (i)(C)(2):

> Whether a specification shows that applicant was in possession of the claimed invention is not a single, simple determination, but rather is a factual determination reached by considering a number of factors. Factors to be considered in determining whether there is sufficient evidence of possession include the level of skill and knowledge in the art, partial structure, physical and/or chemical properties, functional characteristics alone or coupled with a known or disclosed correlation between structure and function, and the method of making the claimed invention. Disclosure of any combination of such identifying characteristics that distinguish the claimed invention from other materials and would lead one of skill in the art to the conclusion that the applicant was in possession of the claimed species is sufficient.

Applicants have satisfied the written description requirement because they have disclosed a combination of identifying characteristics sufficient to distinguish the claimed invention from other materials. *See Amgen Inc. v. Hoechst Marion Roussel, Inc.*, 314 F.3d 1313 (Fed. Cir. 2003). Specifically, Applicants have disclosed structure, physical and/or chemical properties, functional characteristics and a method of making the claimed invention.

First, Applicants have disclosed structure by disclosing the nucleic and amino acid sequences of PRO357, SEQ ID NOS: 68 and 69. Further, those of skill in the art, reading the specification would appreciate that the invention of SEQ ID NO:69 was not limited to only this sequence, but that the inventors contemplated and described a genus of sequences with at least 95% sequence identity to SEQ ID NO: 69. For example, at pages 60-61 of the specification, Applicants disclose methods of making substitutions, as well as substitutions themselves, which could be used to obtain an amino acid sequence variant of the claimed invention, that is one that shares at least 95% sequence identity with SEQ ID NO:69 and that maintains the characteristic of

being amplified in lung or colon tumors. The currently rejected claims are directed to polypeptides such as these.

In addition to describing the structure of the sequence of SEQ ID NO:69, at page 58, lines 1-12, and at page 107, lines 13-17 of the specification, Applicants have disclosed physical and chemical features of SEQ ID NO: 69, which would be common to amino acids that share at least 95% sequence identity with SEQ ID NO: 69. In addition, Figure 26 discloses further features of the encoded polypeptide, which would likely be common to all polypeptides encoded by the claimed genus. Even further, at page 59, Applicants describe variant sequences and explain that:

> Optionally the variation is by substitution of at least one amino acid with any other amino acid in one or more of the domains of the PRO. Guidance in determining which amino acid residue may be inserted, substituted or deleted without adversely affecting the desired activity may be found by comparing the sequence of the PRO with that of homologous known protein molecules and minimizing the number of amino acid sequence changes made in regions of high homology....The variation allowed may be determined by systemically making insertions, deletions or substitutions of amino acids in the sequence and testing the resulting variants for activity exhibited by the full-length or mature native sequence.

Claims 25-36 also require that the claimed polypeptide variants have the characteristic of being encoded by a nucleic acid that is amplified in lung or colon tumors. The Examiner disagrees that being "encoded by a nucleic acid that is amplified in lung or colon tumors" is a "function" of the claimed polypeptide. In any event, as discussed above, a claim to a genus may be adequately described by description of *characteristics* that are common to the genus and allow those of skill in the art to determine whether a particular species falls within the scope of the genus. MPEP § 2163. Being "encoded by a nucleic acid that is amplified in lung or colon tumors" is one characteristic which distinguishes members of the claimed genus from other polypeptides.

As discussed previously, in addition to describing structure, physical and chemical properties and characteristics of the claimed nucleic acids, Applicants also have disclosed how to "make" the claimed invention. Specifically, as discussed previously, at

pages 122-137, Applicants disclose an assay for identifying and isolating the nucleic acids of the claimed invention. More specifically, at pages 120-124 of the specification, Applicants teach that SEQ ID NO: 68, which encodes SEQ ID NO:69, may be isolated from lung or colon tumors. One of skill in the art would appreciate that polypeptides that are at least 95% identical to SEQ ID NO:69 may also be encoded by DNA isolated from lung or colon tumors.

At pages 23-29, and at Table 1, pages 34-54, the specification teaches one of ordinary skill in the art how to determine whether a particular sequence is 95-99% identical to a sequence such as SEQ ID NO: 49. Pages 124-137 of the specification teach one of skill in the art a method for assaying to determine whether a particular sequence has the characteristic of being amplified in lung or colon tumors.

Thus, based on the above combination of described factors, Applicants have demonstrated possession of the claimed invention and provided an adequate written description of the invention. Therefore, Applicants respectfully request that the Examiner withdraw this ground of rejection.

Additionally, Applicants maintain, as previously argued, that the claimed invention satisfies the written description requirement under the analysis of Examples 13 and 14 of the Training Materials which accompany the Written Description Guidelines. The Examiner does not address Applicants' arguments based on Example 13. Specifically, Applicants previously argued that Example 13 explains what is lacking in a description for a claim to a genus of variant proteins. In particular, as stated previously, according to Example 13, a claim to "[a]n isolated variant of the protein of claim 1" is not adequately described if: (1) the specification and claim do not indicate what distinguishing attributes are shard by members of the claimed genus; (2) the specification and claim do not place any limit on the number of amino acid substitutions, deletions, insertions and/or addition; and (3) the specification and claim fail to disclose structural features that could distinguish compounds in the genus from those outside the genus.

Claims 25 and 26 are adequately described under this analysis because, as discussed above, the specification and the claim do indicate distinguishing attributes that are shared by members of the claimed genus, for example being encoded by a nucleic acid that is amplified in lung or colon tumors; the present specification and claims limit the number of substitutions, deletions, insertions, and/or additions by requiring all sequences within the genus to be 95-99% identical to SEQ ID NO:69; and the present specification at page 107 and Figure 26 discloses several structural features common to species falling within the claimed genus. Hence, Example 13 further evidences that the present invention is adequately described.

The Examiner maintains that, contrary to Applicants' assertion, the present claims are not analogous to Example 14 of the Training Materials. Specifically, the Examiner alleges that overexpression in lung and colon tumors is only the "function" of SEQ ID NO:68, and argues that the specification does not disclose any other polypeptides or how one would find a polypeptide that is 95-99% identical that would be encoded by a nucleic acid that is amplified in lung or colon tumors. The Examiner further contends that although the specification teaches a method to assay for expression, that method requires the use of the nucleic acid of the PRO or use of an antibody to the protein of SEQ ID NO:69. Finally, the Examiner argues that the specification does not teach how one would or could find any other polypeptide that is 95-99% identical to SEQ ID NO:69, which is encoded by a nucleic acid that is amplified in lung or colon tumors; nor, the Examiner contends, does the specification teach which regions or parts of the nucleic acid of SEQ ID NO:68 would be used to find such.

Applicants maintain their previous arguments in response to the Examiner's rejections based on Example 14. Additionally, Applicants assert that the gene amplification assay described at pages 119-137 of the specification teaches one of skill in the art how to find a polypeptide that is 95-99% identical to SEQ ID NO: 69, which is encoded by a nucleic acid that is amplified in lung or colon tumors. These pages also teach the regions or parts of the nucleic acid of SEQ ID NO: 68 that could be used to find such

16

polypeptides. Specifically, at page 119, Applicants teach that genomic DNA encoding polypeptides claimed in the present invention can be isolated from lung or colon cancer. Applicants teach at pages 119-120 that the 5' nuclease and real-time PCR reactions can be run using these genomic DNA sequences to determine what genes are potentially amplified. Applicants further explain that the results of the 5' nuclease and real-time PCR assays are quantitated using primers and TaqMan fluorescent probes derived from portions of the genomic DNA that are most likely to contain unique nucleic acid sequences and which are least likely to have spliced out introns, for example, 3'-untranslated regions. Applicants describe the process for preparing DNA to be run in these assays at pages 122-123 of the specification. At pages 123-124, Applications describe isolation of the DNA to be used in these assays. Applicants further describe quantitation of the DNA at page 124. Pages 134-137 of the specification describe framework and epicenter mapping for PRO357 DNA. One of ordinary skill in the art will appreciate that sequences identified using the techniques described at pages 119-137 of the specification may be within the scope of the invention. Those sequences can be confirmed to be within the scope of the present invention by performing the sequence identity analysis described at pages 23-29 and Table 1. Thus, for these reasons in additions to the reasons previously argued, Applicants maintain that the present claims are analogous to those found in Example 14.

In any event, regardless of the similarity between the present claims and those found in Example 14, for all the reasons discussed above, Applicants have satisfied the written description requirement of 35 U.S.C § 112, ¶ 1 and respectfully request this ground of rejection be withdrawn.

Enablement:

The Examiner has also maintained his rejection of Claims 25-36 under 35 U.S.C. § 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to enable one skilled in the art to make and/or use the invention.

Specifically, the Examiner notes that Applicants have previously presented an enablement argument based on the Wands factors. The Examiner disagrees with several portions of the analysis presented by Applicants.

First, the Examiner notes that Applicants argue that "even though Applicants do not specifically state which portions of the disclosed wild type sequence might be altered yet lead to a functional polypeptide, obtaining such sequence variant is not unpredictable and the specification teaches mutagenesis and the PRO357 sequence possesses significant homology to the acid labile subunit of insulin-growth factor and therefore one would compare the claimed polypeptide sequence to the acid labile subunit and minimize amino acid changes in regions of high homology ( )." Office Action at page 6. In response to this argument, the Examiner asserts that "the acid labile subunit is not over expressed in lung or colon tumor and as such why would one look to this sequence."

As Applicants previously explained, Applicants disclose at page 59, lines 24-27 of the specification that "[g]uidance in determining which amino acid residues may be inserted, substituted or deleted without adversely affecting the desired activity may be found by comparing the sequence of the PRO with that of known homologous protein molecules and minimizing the number of amino acid sequence changes made in regions of high homology." Applicants disclose at pages 6, 7, and 107 that portions of the amino acid sequence of the full-length of PRO357 possess significant homology to the acid labile subunit of insulin-growth factor. Therefore, one of ordinary skill in the art, reading the disclosure, would know to compare the claimed polypeptide sequence with the sequence for the acid labile subunit and minimize amino acid changes in regions of high homology between the sequences. Even though the acid labile subunit might not be encoded by a nucleic acid that is amplified in lung or colon tumors, it is still a protein with a specific structure. Those of skill in the art will appreciate that those portions of PRO357 possessing significant homology with the acid labile subunit likely encode structural features of PRO357. Therefore, it would be preferred to not introduce sequence alterations in the region encoding structure. Thus, this explains why one

18

would look to the acid labile subunit sequence when determining what portions of SEQ ID NO:68 or 69 to alter.

The Examiner further contends that Applicants have not addressed the unpredictability in the art as exemplified in the references cited by the Examiner. Specifically, the Examiner cites, Burgess et al., Lazar et al., Schwartz et al., and Lin et al. for the proposition that "even a single modification or substitution in a protein sequence can alter the proteins (sic) function." Applicants previously responded that they did not disagree with the Examiner's assertion but rather agree that a single amino acid modification might result in a significantly altered protein. However, Applicants also responded that although a single modification might result in a significant change, the specification provides one of ordinary skill in the art with ample guidance for selecting sequence modifications that will conserve the function of the encoded polypeptide. For example, at pages 60-62 of the specification, Applicants discuss conservative substitutions that might be used in modifying a sequence that will maintain function following modification. At page 59 of the specification, Applicants teach that modifications to any sequence are preferably not made in regions encoding structure or regions of high homology with other known proteins.

None of the references cited by the Examiner contradict the guidance provided by Applicants in the specification. For example, Burgess et al. examine the effects of using site-directed mutagenesis to change a lysine residue to a glutamic acid residue in a fibroblast. Burgess *et al.,*1990. Possible Dissociation of the Heparin-binding and Mitogenic Activities of Heparin-binding (Acidic Fibroblast) Growth Factor-1 from Its Receptor-binding Activities by Site-directed Mutagenesis of a Single Lysine Residue. *Journal of Cell Biology.* 111:2129-2138. Changing lysine to glutamic acid is not a conservative substitution. As set forth at page 61 of the specification, lysine may be conservatively substituted with arg, gln, and asn. Page 61 further sets forth that lysine is a basic residue while glutamic acid is an acidic residue. Thus, any unpredictability regarding the function of the polypeptide modified in Burgess can be attributed to the modification being a non-conservative modification.

Lazar describes the effects of altering two amino acids that are conserved in the family of EFG-like peptides. Lazar, *et al.,* "Transforming Growth Factor α: Mutation of Aspartic Acid 47 and Leucine 48 Results in Different Biological Activities." *Molecular and Cellular Biology.* 1988. 8(3):1247-52. This teaching of Lazar goes against the teaching of the specification. Specifically, as discussed above, Applicants teach at page 59 of the specification that the sequences of known homologous proteins should be compared with the PRO sequence. This comparison reveals regions of high homology, which are generally conserved sequences. Applicants teach that it is preferable to not introduce modifications in such regions of a sequence, or at least to minimize alterations in such regions. Thus, predictability is increased if modifications are not made in conserved regions of the sequence, which is consistent with the teachings of Lazar.

In Schwartz *et al.,* aspartic acid was substituted for histidine. *See* Schwartz *et al.,* "A superactive insulin: [B10-Aspartic acid]insulin(human)." *PNAS.* 1987. 84:6408-6411. According to page 61 of the specification, this is not a conservative substitution. Conservative substitutions for histidine might include asn, gln, lys, or arg. Therefore, as with regard to Burgess, any art unpredictability taught by this reference can be attributed to the modification being a non-conservative modification.

Lin *et al.* teach that the "amino-terminal histidyl residue in glucagon plays an important but *not obligatory* role in the expression of hormone action and contributes to a significant extent in the recognition process." Lin *et al.,* "Structure-function relationships in glucagons: properties of highly purified des-His-1, monoiodo-, and (des-Asn-28, Thr-29) (homoserine lactone-27)-glucagon." *Biochem.* 1975. 14(8):1559-63 (abstract, emphasis added). Thus, Lin teaches that removal of the amino terminal histidine decreases activity, but it does not teach that removing the histidine residue renders the protein non-functional.

The Examiner also previously cited several abstracts as support for the proposition that "those of skill in the art recognize that expression of mRNA, specific for a tissue type,

20

does not necessarily correlate nor predict equivalent levels of polypeptide expression."
Specifically, the Examiner relies on abstracts of Fu, Powell, Vallejo, and Jang as
evidence that expression of mRNA, specific for a tissue type does not necessarily
correlate or predict equivalent levels of polypeptide expression.

Applicants respectfully disagree with the Examiner's reliance on these abstracts. Only
Jang *et al.* examines a system involving cancerous cells. Specifically, Jang *et al.*
studies whether an increase in metastatic ability could be explained by changes in the
expression of a number of different metastasis-related genes. Jang *et al.,* "An
examination of the effects of hypoxia, acidosis, and glucose starvation on the
expression of metastasis-associated genes in murine tumor cells." *Clin. Exp.*
Metastasis. 1997. 15(5):469-83. Jang *et al.* reported that no overall correlation was
observed but does not conclude that there is no correlation between protein expression
levels and gene amplification levels. Rather, Jang *et al.* conclude that further studies
are required to establish whether changes in protein levels track changes in mRNA
levels for the specific genes examined.

None of Fu, Powell, or Vallejo teach an apt system for comparing against Applicants'
disclosure. Specifically, none of these references discuss a cancer gene that is
amplified and whether that amplification correlates with protein overexpression. More
specifically, as previously explained, Fu *et al.* is not an apt system with which to
compare Applicants' disclosure. Fu does not demonstrate amplification of any gene but
instead studies expression of the p53 gene, which is a tumor suppressor gene. Fu *et*
al., EMBO Journal, 1996. 15:4392-4401. In contrast, the gene amplification associated
with Applicants' invention is a mechanism of activation of oncogenes.

Similarly, Powell *et al.,* examine a gene that is constitutively expressed in human liver
(Powell *et al.,* "Expression of cytochrome P4502E1 in human liver: assessment of
mRNA, genotype and phenotype." *Pharmacogenetics.* 1998. 8(5):411-21), while
Vallejo examines amplification and expression of two transcription factors, one of which
is encoded by nuclear DNA while the other is encoded by mitochondrial DNA. Vallejo *et*

21

al., "Evidence of tissue-specific, post-transcriptional regulation of NRF-2 expression."
Biochimie. 2000. 82(12):1129-33. Vallejo *et al.* report that although no correlation was
observed between the nuclear transcription factor mRNA and protein levels, correlation
was observed with the mitochondrial transcription factor mRNA and protein levels.
Thus, none of these references contradict or outweigh the evidence discussed above, at
pages 5-8, which demonstrates that there is a reasonable correlation between transcript
levels and protein expression.

Noting that the specification only discloses SEQ ID NO:68 as being amplified, the
Examiner maintains that "[a]lthough one may be able to make a polypeptide that is 95-
99% identical to SEQ ID NO: 69, it would be unpredictable which sequence would
encode a polypeptide that is amplified in lung or colon tumor. . . it would be undue
experimentation to determine the myriad of polypeptides that are 95-99% (identical) to
SEQ ID NO:69 which are encoded by nucleic acids that are amplified in lung and colon
tumors."

Applicants respectfully disagree. As discussed previously, the specification provides
significant guidance (*i.e.* no variation in sequences of homology; variation can be
conservative; variation can be in region not encoding structure, *etc*) to be used in
determining what variations could be introduced into a PRO357 sequence yet the
sequence would continue to encode a nucleic acid that is amplified in lung or colon
tumors. Additionally, as discussed above, Applicants disclose a gene amplification
assay that can be used to test the ability of any variant sequence to encode a nucleic
acid that is amplified in lung or colon tumors.

However, the Examiner argues that the disclosures referred to above, in combination
with the characteristics of SEQ ID NOS:68 and 69 are not enough to enable the present
invention because "[t]he specification does not teach how to use such polypeptides or if
the nucleic acid is amplified in tumor." Applicants respectfully disagree and direct the
Examiner's attention to the utility argument set forth at pages 5-8 of this response.

Thus, for all these reasons, Applicants respectfully submit that the specification satisfies the enablement requirement of 35 U.S.C. § 112, ¶ 1 and respectfully request this ground of rejection be withdrawn.
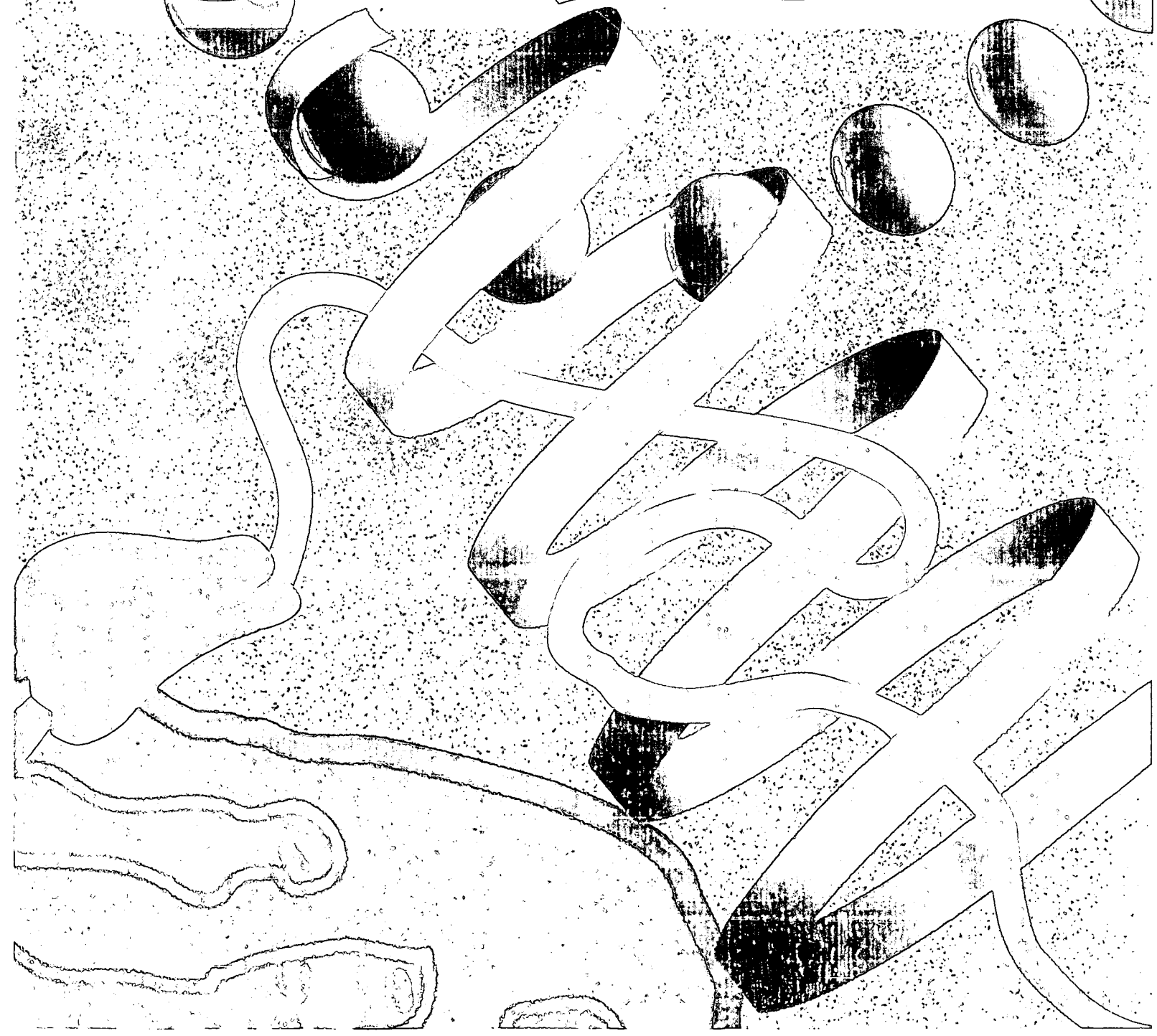
Conclusion

Applicants believe that currently pending Claims 25-36 are patentable.
Applicants respectfully request the Examiner grant allowance of this application. The
Examiner is invited to contact the undersigned attorney for Applicants via telephone if
such communication would expedite the prosecution this application.

Respectfully submitted,

C. Noel Kaman
Registration No. 51,857
Attorney for Applicant

BRINKS HOFER GILSON & LIONE
P.O. BOX 10395
CHICAGO, ILLINOIS 60610
(312) 321-4200

BENJAMIN LEWIN

GENES V

The assembly line for protein synthesis

The **central dogma** defines the paradigm of molecular biology: genes are perpetuated as sequences of nucleic acid, but function by being expressed in the form of proteins.

Three types of processes are responsible for the inheritance of genetic information and for its conversion from one form to another:

♦ Information is *perpetuated* by **replication**; a double-stranded nucleic acid is duplicated to give identical copies.

♦ Information is *expressed* by a two stage process.

♦ **Transcription** generates a single-stranded RNA identical in sequence with one of the strands of the duplex DNA.

♦ **Translation** converts the nucleotide sequence of the RNA into the sequence of amino acids comprising a protein.

The breaking of the genetic code showed that genetic information is stored in the form of nucleotide triplets (codons), but did not reveal *how* each codon specifies its corresponding amino acid. The concept that there must be a code evolved together with the idea that the process of translation must involve a **template** that is separate from the DNA. Because the genetic material in the nucleus is physically separated from the site of protein synthesis in the cytoplasm of a eukaryotic cell, it was

clear that the DNA could not *itself* be translated into protein.

The template is generated by transcription, in the form of a **messenger RNA** (abbreviated **mRNA**) that is identical to one strand of the DNA duplex (see Figure 6.16). We might think of the cell as keeping a 'master set' of sequences in the nucleus, while a 'working set' consists of cytoplasmic mRNA copies of the sequences that are to be expressed.

We distinguish the two strands of DNA as follows:

♦ The DNA strand that bears the *same* sequence as the mRNA (except for possessing T instead of U) is called the **coding strand** or **sense strand**.

♦ The other strand of DNA, which directs synthesis of the mRNA via complementary base pairing, is called the **template strand** or **antisense strand**. (We see later that 'antisense' is used as a general term to describe a sequence of DNA or RNA that is complementary to mRNA.)

Since the genetic code is actually *read* on the mRNA, usually it is described in terms of the four bases present in RNA: U, C, A, and G.

The use of the term *messenger* RNA reflects its ability (in eukaryotes) to move from the nucleus where it is synthesized to the cytoplasm where it functions. Translation of mRNA into protein is accomplished by reading the genetic code: each triplet of nucleotides is converted into one amino acid. Thus 'translation' describes the step at which

Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas*

Torben F. Ørntoft‡§, Thomas Thykjaer¶, Frederic M. Waldman‖, Hans Wolf**, and Julio E. Celis‡‡

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ($p < 0.015$) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ($p < 0.005$) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. *Molecular & Cellular Proteomics 1:37–45, 2002.*

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, *cyclin d1*, *ems1*, and N-myc (3–5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of c-myc copy number increase was observed without concomitant c-myc protein overexpression (6).

In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)[1] have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q−, 11p−, 1q+, 11q13+, 17q+, and 20q+ (7–12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

EXPERIMENTAL PROCEDURES

Material—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary),

[1] The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.
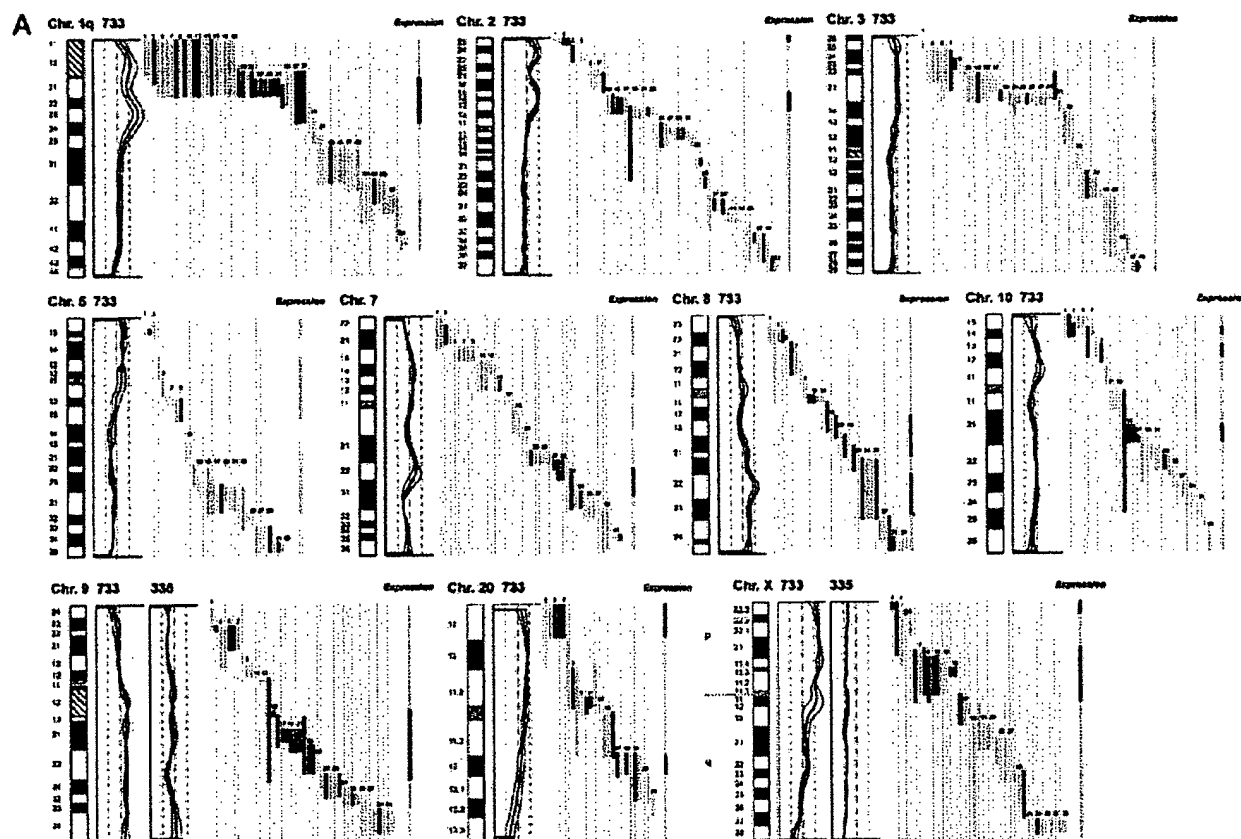
FIG. 1. **DNA copy number and mRNA expression level.** Shown from *left* to *right* are chromosome (*Chr.*), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. *A*, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. *B*, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (*left*). The *bold curve* in the ratio profile represents a mean of four chromosomes and is surrounded by *thin curves* indicating one standard deviation. The *central vertical line (broken)* indicates a ratio value of 1 (no change), and the *vertical lines* next to it (*dotted*) indicate a ratio of 0.5 (*left*) and 2.0 (*right*). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the *right* of the invasive tumor profile. The *colored bars* represents one gene each, identified by the running *numbers* above the *bars* (the name of the gene can be seen at www.MDL.DK/sdata.html). The *bars* indicate the purported location of the gene, and the *colors* indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart; >2-fold increase (*black*), >2-fold decrease (*blue*), no significant change (*orange*). The *bar* to the *far right*, entitled *Expression* shows the resulting change in expression along the chromosome; the *colors* indicate that at least half of the genes were up-regulated (*black*), at least half of the genes down-regulated (*blue*), or more than half of the genes are unchanged (*orange*). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

mRNA Preparation—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at −80 °C. Total RNA was isolated using the RNAzol B RNA isolation method (WAK-Chemie Medical GMBH). poly(A)$^+$ RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

cRNA Preparation—1 μg of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript® choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the ME-GAscrip® *in vitro* transcription kit (Ambion). Biotin-labeled CTP and

UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

Array Hybridization and Scanning—Array hybridization and scanning was modified from a previous method (13). 10 μg of cRNA was fragmented at 94 °C for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6× SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to 95 °C for 5 min, subsequently cooled to 40 °C, and loaded onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at 40 °C at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6× SSPE-T at 25 °C followed by 4 washes in 0.5× SSPE-T at 50 °C. The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10 μg/ml (Molecular Probes) in 6× SSPE-T
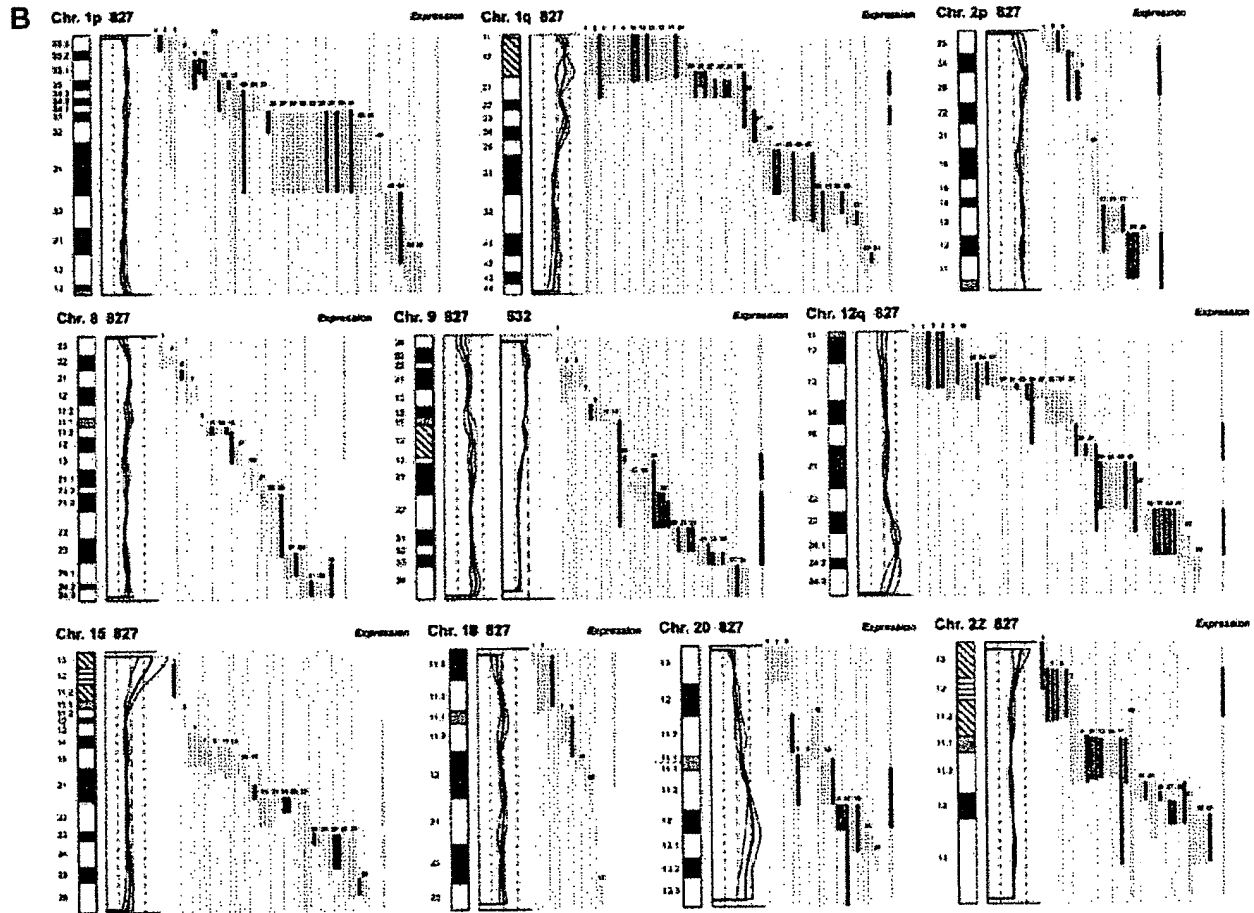
FIG. 1—continued

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

Microsatellite Analysis—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of www.ncbi.nlm.nih.gov/genemap98, and primer sequences were obtained from the genome data base at www.gdb.org. DNA was extracted from tumor and blood and amplified by PCR in a volume of 20 μl for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

Proteomic Analysis—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at −20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see biobase.dk/cgi-bin/cells.

CGH—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-labeled reference DNA (200 ng), and human Cot-1 DNA (20 μg) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15 μg/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

RESULTS

Comparative Genomic Hybridization—The CGH analysis identified a number of chromosomal gains and losses in the

TABLE I

Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (if CGH alteration — what expression ratio was found); bottom, altered expression used as independent variable (if expression alteration — what CGH deviation was found).

CGH alterations	Tumor 733 vs. 335 Expression change clusters	Concordance	CGH alterations	Tumor 827 vs. 532 Expression change clusters	Concordance
13 Gain	10 Up-regulation 0 Down-regulation 3 No change	77%	10 Gain	8 Up-regulation 0 Down-regulation 2 No change	80%
10 Loss	1 Up-regulation 5 Down-regulation 4 No change	50%	12 Loss	3 Up-regulation 2 Down regulation 7 No change	17%

Expression change clusters	Tumor 733 vs. 335 CGH alterations	Concordance	Expression change clusters	Tumor 827 vs. 532 CGH alterations	Concordance
16 Up-regulation	11 Gain 2 Loss 3 No change	69%	17 Up-regulation	10 Gain 5 Loss 2 No change	59%
21 Down-regulation	1 Gain 8 Loss 12 No change	38%	9 Down-regulation	0 Gain 3 Loss 6 No change	33%
15 No change	3 Gain 3 Loss 9 No change	60%	21 No change	1 Gain 3 Loss 17 No change	81%

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p−, 9q22-q33−, and X−, and 7+, 9q−, and Y−, respectively. Both invasive tumors showed changes (1q22-24+, 2q14.1-qter−, 3q12-q13.3−, 6q12-q22−, 9q34+, 11q12-q13+, 17+, and 20q11.2-q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1A) and 20q12 in TCC 827 (Fig. 1B).

mRNA Expression in Relation to DNA Copy Number—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive *versus* the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger

than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21-q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, top). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, top). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, bottom). Furthermore, as a control we looked at areas with no alter-
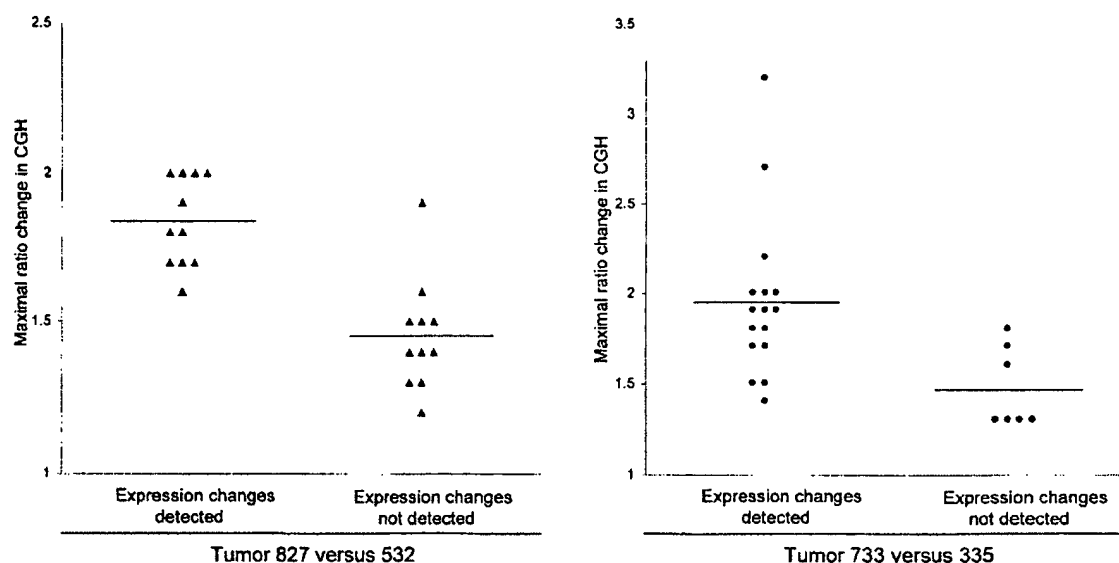
FIG. 2. **Correlation between maximum CGH aberration and the ability to detect expression change by oligonucleotide array monitoring.** The aberration is shown as a numerical -fold change in ratio between invasive tumors 827 (▲) and 733 (♦) and their non-invasive counterparts 532 and 335. The expression change was taken from the *Expression* line to the *right* in Fig. 1, which depicts the resulting expression change for a given chromosomal region. At least half of the mRNAs from a given region have to be either up- or down-regulated to be scored as an expression change. All chromosomal arms in which the CGH ratio plus or minus one standard deviation was outside the ratio value of one were included.

ation in expression. No alteration was detected by CGH in most of these areas (TCC 733, 60% and TCC 827, 81%; see Table I, *bottom*). Because the ability to observe reduced or increased mRNA expression clustering to a certain chromosomal area clearly reflected the extent of copy number changes, we plotted the maximum CGH aberrations in the regions showing CGH changes against the ability to detect a change in mRNA expression as monitored by the oligonucleotide arrays (Fig. 2). For both tumors TCC 733 ($p < 0.015$) and TCC 827 ($p < 0.00003$) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology (Fig. 2). Similar data were obtained when areas with altered expression were used as independent variables. These areas correlated best with CGH when the CGH ratio deviated 1.6- to 2.0-fold (Table I, *bottom*) but mostly did not at lower CGH deviations. These data probably reflect that loss of an allele may only lead to a 50% reduction in expression level, which is at the cut-off point for detection of expression alterations. Gain of chromosomal material can occur to a much larger extent.

Microsatellite-based Detection of Minor Areas of Losses—In TCC 733, several chromosomal areas exhibiting DNA amplification were preceded or followed by areas with a normal CGH but reduced mRNA expression (see Fig. 1, TCC 733 chromosome 1q32, 2p21, and 7q21 and q32, 9q34, and 10q22). To determine whether these results were because of undetected loss of chromosomal material in these regions or

because of other non-structural mechanisms regulating transcription, we examined two microsatellites positioned at chromosome 1q25–32 and two at chromosome 2p22. Loss of heterozygosity (LOH) was found at both 1q25 and at 2p22 indicating that minor deleted areas were not detected with the resolution of CGH (Fig. 3). Additionally, chromosome 2p in TCC 733 showed a CGH pattern of gain/no change/gain of DNA that correlated with transcript increase/decrease/increase. Thus, for the areas showing increased expression there was a correlation with the DNA copy number alterations (Fig. 1A). As indicated above, the mRNA decrease observed in the middle of the chromosomal gain was because of LOH, implying that one of the mechanisms for mRNA down-regulation may be regions that have undergone smaller losses of chromosomal material. However, this cannot be detected with the resolution of the CGH method.

In both TCC 733 and TCC 827, the telomeric end of chromosome 11p showed a normal ratio in the CGH analysis; however, clusters of five and three genes, respectively, lost their expression. Two microsatellites (D11S1760, D11S922) positioned close to MUC2, IGF2, and cathepsin D indicated LOH as the most likely mechanism behind the loss of expression (data not shown).

A reduced expression of mRNA observed in TCC 733 at chromosomes 3q24, 11p11, 12p12.2, 12q21.1, and 16q24 and in TCC 827 at chromosome 11p15.5, 12p11, 15q11.2, and 18q12 was also examined for chromosomal losses using microsatellites positioned as close as possible to the gene loci
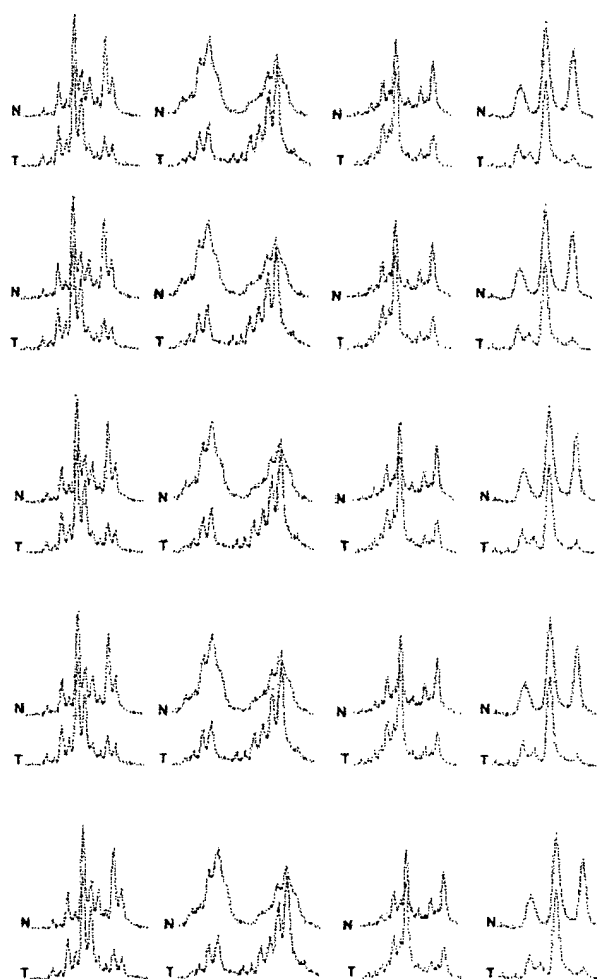
FIG. 3. **Microsatellite analysis of loss of heterozygosity.** Tumor 733 showing loss of heterozygosity at chromosome 1q25, detected (a) by D1S215 close to Hu class I histocompatibility antigen (gene number 38 in Fig. 1), (b) by D1S2735 close to cathepsin E (gene number 41 in Fig. 1), and (c) at chromosome 2p23 by D2S2251 close to general β-spectrin (gene number 11 on Fig. 1) and of (d) tumor 827 showing loss of heterozygosity at chromosome 18q12 by S18S1118 close to mitochondrial 3-oxoacyl-coenzyme A thiolase (gene number 12 in Fig. 1). The *upper curves* show the electropherogram obtained from normal DNA from leukocytes (*N*), and the *lower curves* show the electropherogram from tumor DNA (*T*). In all cases one allele is partially lost in the tumor amplicon.

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

Relation between Changes in mRNA and Protein Levels— 2D-PAGE analysis, in combination with Coomassie Brilliant Blue and/or silver staining, was carried out on all four tumors using fresh biopsy material. 40 well resolved abundant known proteins migrating in areas away from the edges of the pH
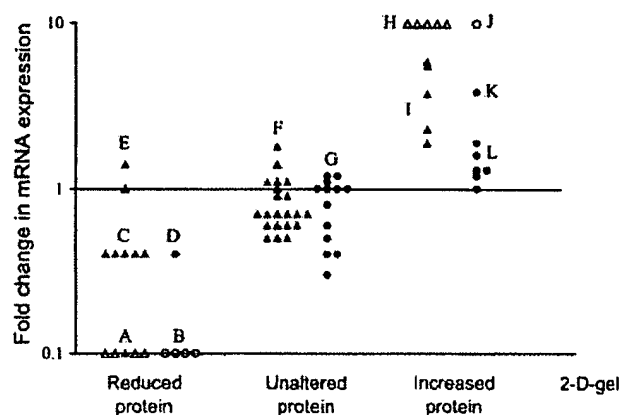

FIG. 4. **Correlation between protein levels as judged by 2D-PAGE and transcript ratio.** For comparison proteins were divided in three groups, unaltered in level or up- or down-regulated (*horizontal axis*). The mRNA ratio as determined by oligonucleotide arrays was plotted for each gene (*vertical axis*). ▲, mRNAs that were scored as present in both tumors used for the ratio calculation; △, mRNAs that were scored as absent in the invasive tumors (along *horizontal axis*) or as absent in non-invasive reference (*top* of figure). Two different scalings were used to exclude scaling as a confounder, TCCs 827 and 532 (▲△) were scaled with background suppression, and TCCs 733 and 335 (●○) were scaled without suppression. Both comparisons showed highly significant (*p* < 0.005) differences in mRNA ratios between the groups. Proteins shown were as follows: Group A (from *left*), phosphoglucomutase 1, glutathione transferase class μ number 4, fatty acid-binding protein homologue, cytokeratin 15, and cytokeratin 13; B (from *left*), fatty acid-binding protein homologue, 28-kDa heat shock protein, cytokeratin 13, and calcyclin; C (from *left*), α-enolase, hnRNP B1, 28-kDa heat shock protein, 14-3-3-ε, and pre-mRNA splicing factor; D, mesothelial keratin K7 (type II); E (from *top*), glutathione S-transferase-π and mesothelial keratin K7 (type II); F (from *top* and *left*), adenylyl cyclase-associated protein, E-cadherin, keratin 19, calgizzarin, phosphoglycerate mutase, annexin IV, cytoskeletal γ-actin, hnRNP A1, integral membrane protein calnexin (IP90), hnRNP H, brain-type clathrin light chain-a, hnRNP F, 70-kDa heat shock protein, heterogeneous nuclear ribonucleoprotein A/B, translationally controlled tumor protein, liver glyceraldehyde-3-phosphate dehydrogenase, keratin 8, aldehyde reductase, and Na,K-ATPase β-1 subunit; G, (from *top* and *left*), TCP20, calgizzarin, 70-kDa heat shock protein, calnexin, hnRNP H, cytokeratin 15, ATP synthase, keratin 19, triosephosphate isomerase, hnRNP F, liver glyceraldehyde-3-phosphatase dehydrogenase, glutathione S-transferase-π, and keratin 8; H (from *left*), plasma gelsolin, autoantigen calreticulin, thioredoxin, and NAD+-dependent 15 hydroxyprostaglandin dehydrogenase; I (from *top*), prolyl 4-hydroxylase β-subunit, cytokeratin 20, cytokeratin 17, prohibition, and fructose 1,6-biphosphatase; J annexin II; K, annexin IV; L (from *top* and *left*), 90-kDa heat shock protein, prolyl 4-hydroxylase β-subunit, α-enolase, GRP 78, cyclophilin, and cofilin.

gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation (*p* < 0.005) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-
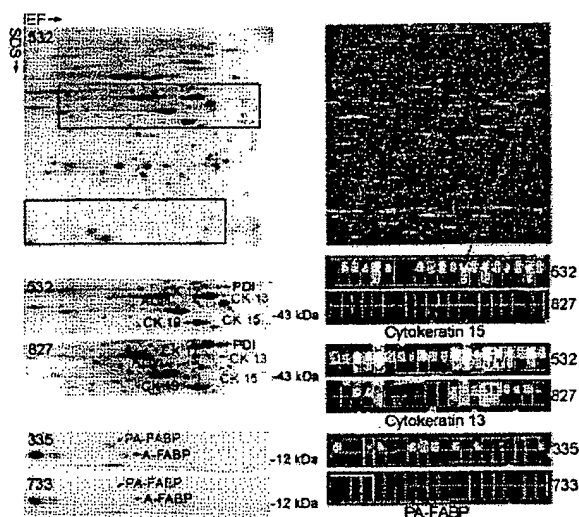
FIG. 5. **Comparison of protein and transcript levels in invasive and non-invasive TCCs.** The *upper part* of the figure shows a 2D gel (*left*) and the oligonucleotide array (*right*) of TCC 532. The *red rectangles* on the *upper gel* highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (*red annotation*). The tile on the array containing probes for cytokeratin 15 is enlarged *below* the array (*red arrow*) from TCC 532 and is compared with TCC 827. The *upper row* of *squares* in each tile corresponds to perfect match probes; the *lower row* corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as *black*, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (6151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the *bottom* of the figure (*left*) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the *right* we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). *IEF,* isoelectric focusing.

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a know chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p < 0.005$) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FBP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

TABLE II
Proteins whose expression level correlates with both mRNA and gene dose changes

Protein	Chromosomal location	Tumor TCC	CGH alteration	Transcript alteration[a]	Protein alteration
Annexin II	1q21	733	Gain	Abs to Pres[a]	Increase
Annexin IV	2p13	733	Gain	3.9-Fold up	Increase
Cytokeratin 17	17q12-q21	827	Gain	3.8-Fold up	Increase
Cytokeratin 20	17q21.1	827	Gain	5.6-Fold up	Increase
(PA-)FABP	8q21.2	827	Loss	10-Fold down	Decrease
FBP1	9q22	827	Gain	2.3-Fold up	Increase
Plasma gelsolin	9q31	827	Gain	Abs to Pres	Increase
Heat shock protein 28	15q12-q13	827	Loss	2.5-Fold up	Decrease
Prohibitin	17q21	827/733	Gain	3.7-/2.5-Fold up[b]	Increase
Prolyl-4-hydroxyl	17q25	827/733	Gain	5.7-/1.6-Fold up	Increase
hnRNPB1	7p15	827	Loss	2.5-Fold down	Decrease

[a] Abs, absent; Pres, present.
[b] In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (*e.g.* in TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; in TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17–19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p−, 9q−, 1q+, Y− (2, 6), and in pT1 tumors, 2q−,11p−, 11q−, 1q+, 5p+, 8q+, 17q+, and 20q+ (2–4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p− and 9q22-q33− and 9q− and Y−, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22–24 amplification (seen in both tumors), 11q14-q22 loss, the latter often linked to 17 q+ (both tumors), and 1q+ and 9p−, often linked to 20q+ and 11 q13+ (both tumors) (7–9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal

arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicate that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (*e.g.* signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker *et al.* (26) in yeast.

Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript. One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

REFERENCES

1. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* 17, 643–649
2. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 23, 41–46
3. de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* 83, 157–161
4. Brungier, P. P., Tamimi, Y., Shuuring, E., and Schalken, J. (1996) Expression of cyclin D1 and EMS1 in bladder tumors; relationship with chromosome 11q13 amplifications. *Oncogene* 12, 1747–1753
5. Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) *myc* gene amplification and expression in primary human neuroblastoma. *Cancer Res.* 50, 1459–1463
6. Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) c-*myc* copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am J. Pathol.* 146, 1131–1139
7. Richter, J., Jiang, F., Gorog, J. P., Sartorious, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* 57, 2860–2864
8. Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* 153, 1615–1621
9. Bruch, J., Wohr, G., Hautmann, R., Mattfeldt, T., Bruderlein, S., Moller, P., Sauter, S., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* 23, 167–174
10. Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* 15, 3555–3560
11. Simon, R., Burger, H., Brinkschmidt, C., Bocker, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* 185, 345–351
12. Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* 110, 87–93
13. Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae. Nat. Biotechnol.* 15, 1359–1367
14. Christensen, M., Sunde, L., Bolund, L., and Orntoft, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* 59, 167–177
15. Celis, J. E., Ostergaard, M., Basse, B., Celis, A., Lauridsen, J. B., Ratz, G. P., Andersen, I., Hein, B., Wolf, H., Orntoft, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* 56, 4782–4790
16. Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994) in *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 3, pp. 222–230, Academic Press, Orlando, FL
17. Ohlsson, R., Tycko, B., and Sapienza, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* 14, 435–438
18. Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the interleukin-2 locus. *Science* 279, 2118–2121
19. Brannan, C. I., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* 9, 164–170
20. Ohlsson, R., Cui, H., He, L., Pfeifer, S., Malmikumpa, H., Jiang, S., Feinberg, A. P., and Hedborg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* 59, 3889–3892
21. Cui, H., Hedborg, F., He, L., Nordenskjold, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* 57, 4469–4473
22. Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* 285, 251–254
23. Tsao, J., Yatabe, Y., Markl, I. D., Hajyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* 29, 26–32
24. Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10632–10636
25. Anderson, L., and Seilhamer, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18, 533–537
26. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934

Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors

Jonathan R. Pollack*††, Therese Sørlie§, Charles M. Perou¶, Christian A. Rees‖**, Stefanie S. Jeffrey††, Per E. Lonning‡‡, Robert Tibshirani§§, David Botstein‖, Anne-Lise Børresen-Dale§, and Patrick O. Brown†¶¶

Departments of *Pathology, ‖Genetics, ††Surgery, §§Health Research and Policy, and ¶¶Biochemistry, and †Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; §Department of Genetics, Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway; ‡‡Department of Medicine (Oncology), Haukeland University Hospital, N-5021 Bergen, Norway; and ¶Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

Contributed by Patrick O. Brown, August 6, 2002

Genomic DNA copy number alterations are key genetic events in the development and progression of human cancers. Here we report a genome-wide microarray comparative genomic hybridization (array CGH) analysis of DNA copy number variation in a series of primary human breast tumors. We have profiled DNA copy number alteration across 6,691 mapped human genes, in 44 predominantly advanced, primary breast tumors and 10 breast cancer cell lines. While the overall patterns of DNA amplification and deletion corroborate previous cytogenetic studies, the high-resolution (gene-by-gene) mapping of amplicon boundaries and the quantitative analysis of amplicon shape provide significant improvement in the localization of candidate oncogenes. Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells. Specifically, we find that 62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels, and that overall, at least 12% of all the variation in gene expression among the breast tumors is directly attributable to underlying variation in gene copy number. These findings provide evidence that widespread DNA copy number alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer.

Conventional cytogenetic techniques, including comparative genomic hybridization (CGH) (1), have led to the identification of a number of recurrent regions of DNA copy number alteration in breast cancer cell lines and tumors (2–4). While some of these regions contain known or candidate oncogenes [e.g., FGFR1 (8p11), MYC (8q24), CCND1 (11q13), ERBB2 (17q12), and ZNF217 (20q13)] and tumor suppressor genes [RB1 (13q14) and TP53 (17p13)], the relevant gene(s) within other regions (e.g., gain of 1q, 8q22, and 17q22–24, and loss of 8p) remain to be identified. A high-resolution genome-wide map, delineating the boundaries of DNA copy number alterations in tumors, should facilitate the localization and identification of oncogenes and tumor suppressor genes in breast cancer. In this study, we have created such a map, using array-based CGH (5–7) to profile DNA copy number alteration in a series of breast cancer cell lines and primary tumors.

An unresolved question is the extent to which the widespread DNA copy number changes that we and others have identified in breast tumors alter expression of genes within involved regions. Because we had measured mRNA levels in parallel in the same samples (8), using the same DNA microarrays, we had an opportunity to explore on a genomic scale the relationship between DNA copy number changes and gene expression. From this analysis, we have identified a significant impact of widespread DNA copy number alteration on the transcriptional programs of breast tumors.

Materials and Methods

Tumors and Cell Lines. Primary breast tumors were predominantly large (>3 cm), intermediate-grade, infiltrating ductal carcinomas, with more than 50% being lymph node positive. The fraction of tumor cells within specimens averaged at least 50%. Details of individual tumors have been published (8, 9), and are summarized in Table 1, which is published as supporting information on the PNAS web site, www.pnas.org. Breast cancer cell lines were obtained from the American Type Culture Collection. Genomic DNA was isolated either using Qiagen genomic DNA columns, or by phenol/chloroform extraction followed by ethanol precipitation.

DNA Labeling and Microarray Hybridizations. Genomic DNA labeling and hybridizations were performed essentially as described in Pollack et al. (7), with slight modifications. Two micrograms of DNA was labeled in a total volume of 50 microliters and the volumes of all reagents were adjusted accordingly. "Test" DNA (from tumors and cell lines) was fluorescently labeled (Cy5) and hybridized to a human cDNA microarray containing 6,691 different mapped human genes (i.e., UniGene clusters). The "reference" (labeled with Cy3) for each hybridization was normal female leukocyte DNA from a single donor. The fabrication of cDNA microarrays and the labeling and hybridization of mRNA samples have been described (8).

Data Analysis and Map Positions. Hybridized arrays were scanned on a GenePix scanner (Axon Instruments, Foster City, CA), and fluorescence ratios (test/reference) calculated using SCANALYZE software (available at http://rana.lbl.gov). Fluorescence ratios were normalized for each array by setting the average log fluorescence ratio for all array elements equal to 0. Measurements with fluorescence intensities more than 20% above background were considered reliable. DNA copy number profiles that deviated significantly from background ratios measured in normal genomic DNA control hybridizations were interpreted as evidence of real DNA copy number alteration (see *Estimating Significance of Altered Fluorescence Ratios* in the supporting information). When indicated, DNA copy number profiles are displayed as a moving average (symmetric 5-nearest neighbors). Map positions for arrayed human cDNAs were assigned by
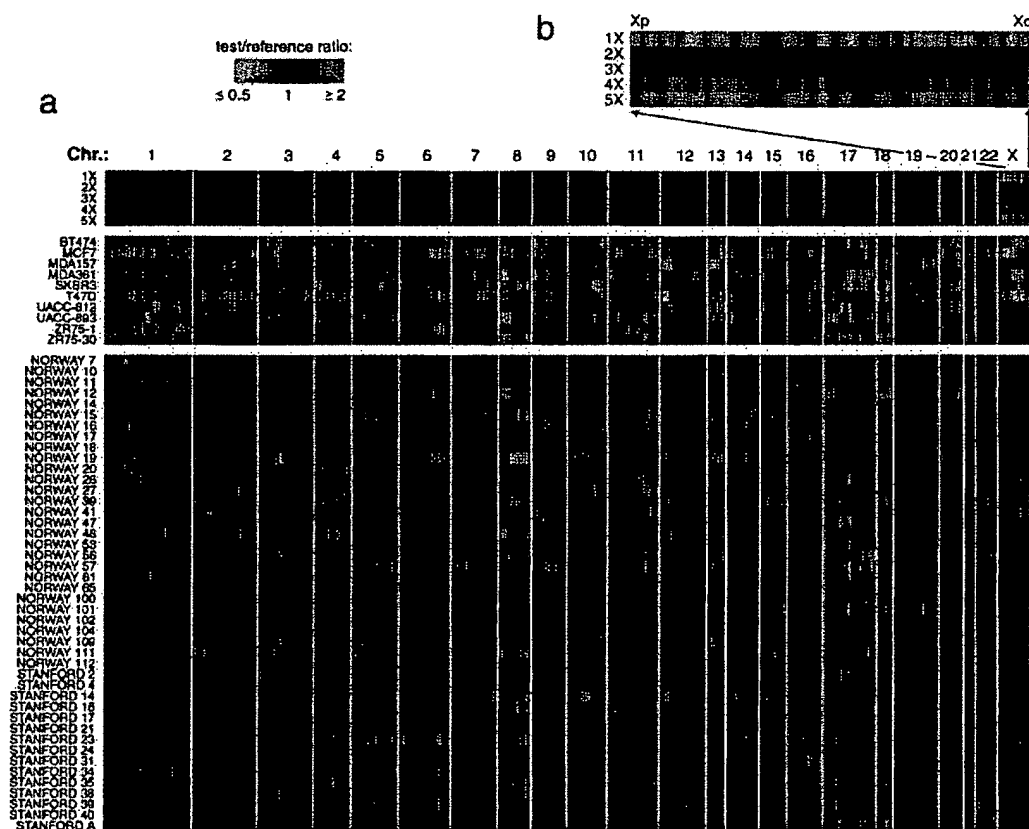
GENETICS

Fig. 1. Genome-wide measurement of DNA copy number alteration by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Each row represents a different cell line or tumor, and each column represents one of 6,691 different mapped human genes present on the microarray, ordered by genome map position from 1pter through Xqter. Moving average (symmetric 5-nearest neighbors) fluorescence ratios (test/reference) are depicted using a log₂-based pseudocolor scale (indicated), such that red luminescence reflects fold-amplification, green luminescence reflects fold-deletion, and black indicates no change (gray indicates poorly measured data). (b) Enlarged view of DNA copy number profiles across the X chromosome, shown for cell lines containing different numbers of X chromosomes.

identifying the starting position of the best and longest match of any DNA sequence represented in the corresponding UniGene cluster (10) against the "Golden Path" genome assembly (http://genome.ucsc.edu/; Oct 7, 2000 Freeze). For UniGene clusters represented by multiple arrayed elements, mean fluorescence ratios (for all elements representing the same UniGene cluster) are reported. For mRNA measurements, fluorescence ratios are "mean-centered" (i.e., reported relative to the mean ratio across the 44 tumor samples). The data set described here can be accessed in its entirety in the supporting information.

Results

We performed CGH on 44 predominantly locally advanced, primary breast tumors and 10 breast cancer cell lines, using cDNA microarrays containing 6,691 different mapped human genes (Fig. 1a; also see Materials and Methods for details of microarray hybridizations). To take full advantage of the improved spatial resolution of array CGH, we ordered (fluorescence ratios for) the 6,691 cDNAs according to the "Golden Path" (http://genome.ucsc.edu/) genome assembly of the draft human genome sequences (11). In so doing, arrayed cDNAs not only themselves represent genes of potential interest (e.g., candidate oncogenes within amplicons), but also provide precise genetic landmarks for chromosomal regions of amplification and

deletion. Parallel analysis of DNA from cell lines containing different numbers of X chromosomes (Fig. 1b), as we did before (7), demonstrated the sensitivity of our method to detect single-copy loss (45, XO), and 1.5- (47,XXX), 2- (48,XXXX), or 2.5-fold (49,XXXXX) gains (also see Fig. 5, which is published as supporting information on the PNAS web site). Fluorescence ratios were linearly proportional to copy number ratios, which were slightly underestimated, in agreement with previous observations (7). Numerous DNA copy number alterations were evident in both the breast cancer cell lines and primary tumors (Fig. 1a), detected in the tumors despite the presence of euploid non-tumor cell types; the magnitudes of the observed changes were generally lower in the tumor samples. DNA copy-number alterations were found in every cancer cell line and tumor, and on every human chromosome in at least one sample. Recurrent regions of DNA copy number gain and loss were readily identifiable. For example, gains within 1q, 8q, 17q, and 20q were observed in a high proportion of breast cancer cell lines/tumors (90%/69%, 100%/47%, 100%/60%, and 90%/44%, respectively), as were losses within 1p, 3p, 8p, and 13q (80%/24%, 80%/22%, 80%/22%, and 70%/18%, respectively), consistent with published cytogenetic studies (refs. 2–4; a complete listing of gains/losses is provided in Tables 2 and 3, which are published as supporting information on the PNAS web site). The total
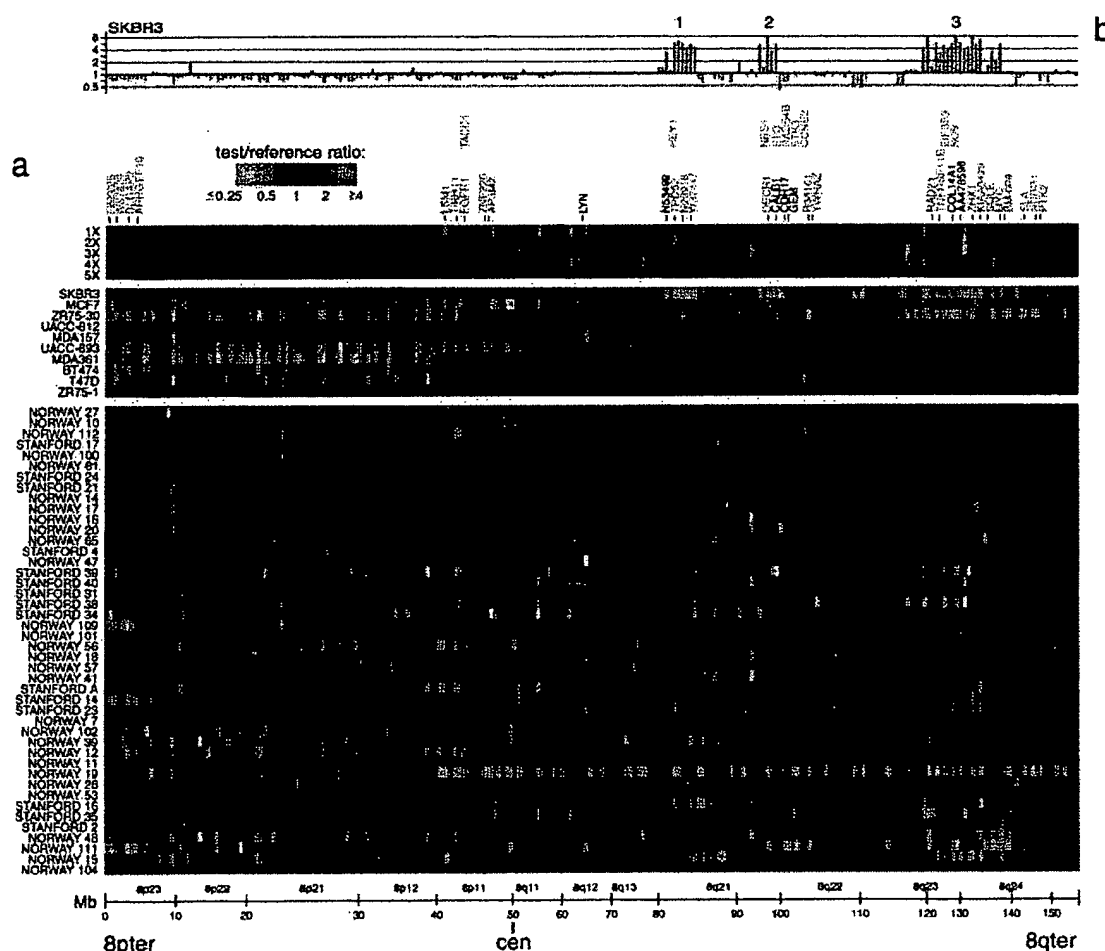
Fig. 2. DNA copy number alteration across chromosome 8 by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering to highlight recurrent copy number changes. The 241 genes present on the microarrays and mapping to chromosome 8 are ordered by position along the chromosome. Fluorescence ratios (test/reference) are depicted by a log$_2$ pseudocolor scale (indicated). Selected genes are indicated with color-coded text (red, increased; green, decreased; black, no change; gray, not well measured) to reflect correspondingly altered mRNA levels (observed in the majority of the subset of samples displaying the DNA copy number change). The map positions for genes of interest that are not represented on the microarray are indicated in the row above those genes represented on the array. (b) Graphical display of DNA copy number profile for breast cancer cell line SKBR3. Fluorescence ratios (tumor/normal) are plotted on a log$_2$ scale for chromosome 8 genes, ordered along the chromosome.

number of genomic alterations (gains and losses) was found to be significantly higher in breast tumors that were high grade ($P = 0.008$), consistent with published CGH data (3), estrogen receptor negative ($P = 0.04$), and harboring TP53 mutations ($P = 0.0006$) (see Table 4, which is published as supporting information on the PNAS web site).

The improved spatial resolution of our array CGH analysis is illustrated for chromosome 8, which displayed extensive DNA copy number alteration in our series. A detailed view of the variation in the copy number of 241 genes mapping to chromosome 8 revealed multiple regions of recurrent amplification; each of these potentially harbors a different known or previously uncharacterized oncogene (Fig. 2a). The complexity of amplicon structure is most easily appreciated in the breast cancer cell line SKBR3. Although a conventional CGH analysis of 8q in SKBR3 identified only two distinct regions of amplification (12), we observed three distinct regions of high-level amplification (labeled 1-3 in Fig. 2b). For each of these regions we can define the

boundaries of the interval recurrently amplified in the tumors we examined; in each case, known or plausible candidate oncogenes can be identified (a description of these regions, as well as the recurrently amplified regions on chromosomes 17 and 20, can be found in Figs. 6 and 7, which are published as supporting information on the PNAS web site).

For a subset of breast cancer cell lines and tumors (4 and 37, respectively), and a subset of arrayed genes (6,095), mRNA levels were quantitatively measured in parallel by using cDNA microarrays (8). The parallel assessment of mRNA levels is useful in the interpretation of DNA copy number changes. For example, the highly amplified genes that are also highly expressed are the strongest candidate oncogenes within an amplicon. Perhaps more significantly, our parallel analysis of DNA copy number changes and mRNA levels provides us the opportunity to assess the global impact of widespread DNA copy number alteration on gene expression in tumor cells.

A strong influence of DNA copy number on gene expression is evident in an examination of the pseudocolor representations
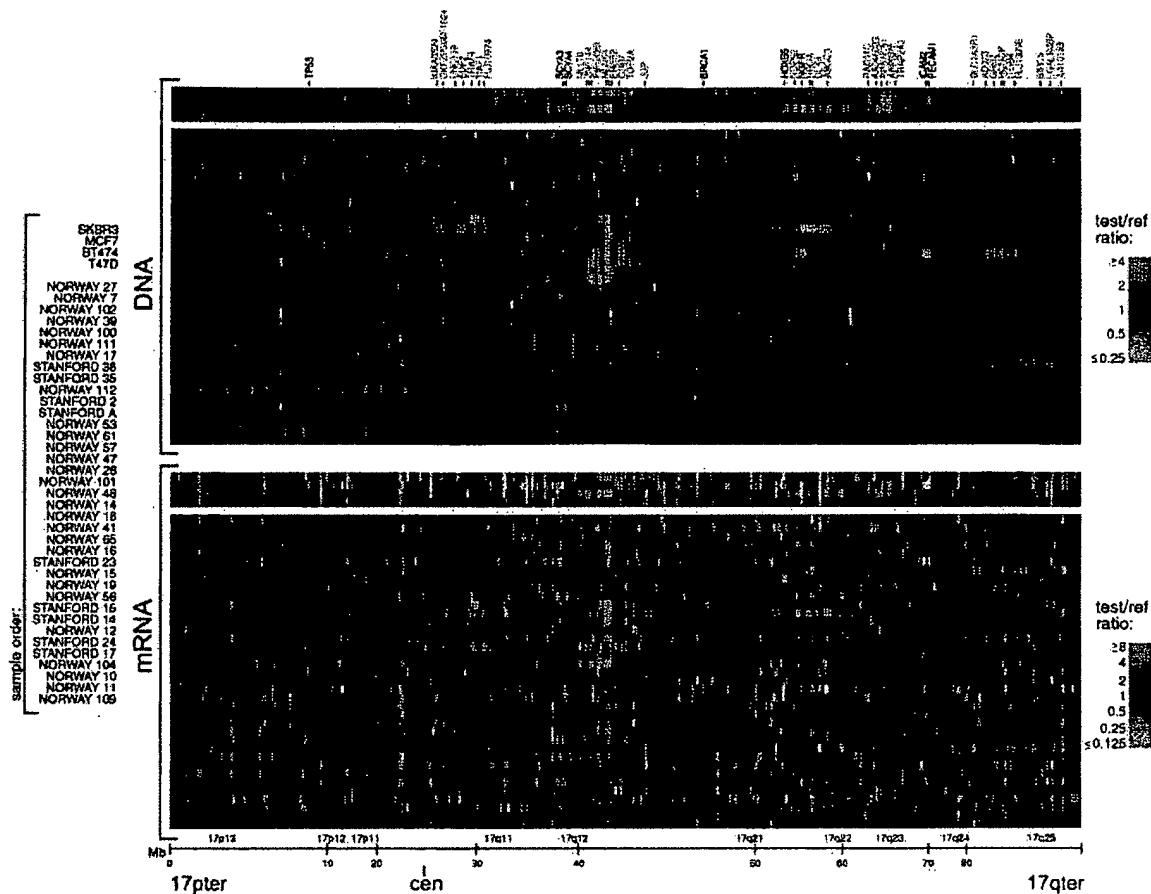
Fig. 3. Concordance between DNA copy number and gene expression across chromosome 17. DNA copy number alteration (*Upper*) and mRNA levels (*Lower*) are illustrated for breast cancer cell lines and tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering (*Upper*), and the identical sample order is maintained (*Lower*). The 354 genes present on the microarrays and mapping to chromosome 17, and for which both DNA copy number and mRNA levels were determined, are ordered by position along the chromosome; selected genes are indicated in color-coded text (see Fig. 2 legend). Fluorescence ratios (test/reference) are depicted by separate $\log_2$ pseudocolor scales (indicated).

of DNA copy number and mRNA levels for genes on chromosome 17 (Fig. 3). The overall patterns of gene amplification and elevated gene expression are quite concordant; i.e., a significant fraction of highly amplified genes appear to be correspondingly highly expressed. The concordance between high-level amplification and increased gene expression is not restricted to chromosome 17. Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios >4, and representing 91 different genes), 62% (representing 54 different genes; see Table 5, which is published as supporting information on the PNAS web site) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios >2), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios >4).

To determine the extent to which DNA deletion and lower-level amplification (in addition to high-level amplification) are also associated with corresponding alterations in mRNA levels, we performed three separate analyses on the complete data set (4 cell lines and 37 tumors, across 6,095 genes). First, we determined the average mRNA levels for each of five classes of genes, representing DNA deletion, no change, and low-, medium-, and high-level amplification (Fig. 4*a*). For both the

breast cancer cell lines and tumors, average mRNA levels tracked with DNA copy number across all five classes, in a statistically significant fashion (*P* values for pair-wise Student's *t* tests comparing adjacent classes: cell lines, 4×10^{-49}, 1×10^{-49}, 5×10^{-5}, 1×10^{-2}; tumors, 1×10^{-43}, 1×10^{-214}, 5×10^{-41}, 1×10^{-4}). A linear regression of the average log(DNA copy number), for each class, against average log(mRNA level) demonstrated that on average, a 2-fold change in DNA copy number was accompanied by 1.4- and 1.5-fold changes in mRNA level for the breast cancer cell lines and tumors, respectively (Fig. 4*a*, regression line not shown). Second, we characterized the distribution of the 6,095 correlations between DNA copy number and mRNA level, each across the 37 tumor samples (Fig. 4*b*). The distribution of correlations forms a normal-shaped curve, but with the peak markedly shifted in the positive direction from zero. This shift is statistically significant, as evidenced in a plot of observed vs. expected correlations (Fig. 4*c*), and reflects a pervasive global influence of DNA copy number alterations on gene expression. Notably, the highest correlations between DNA copy number and mRNA level (the right tail of the distribution in Fig. 4*b*) comprise both amplified and deleted genes (data not shown). Third, we used a linear regression model to estimate the fraction of all variation measured in mRNA levels among the 37
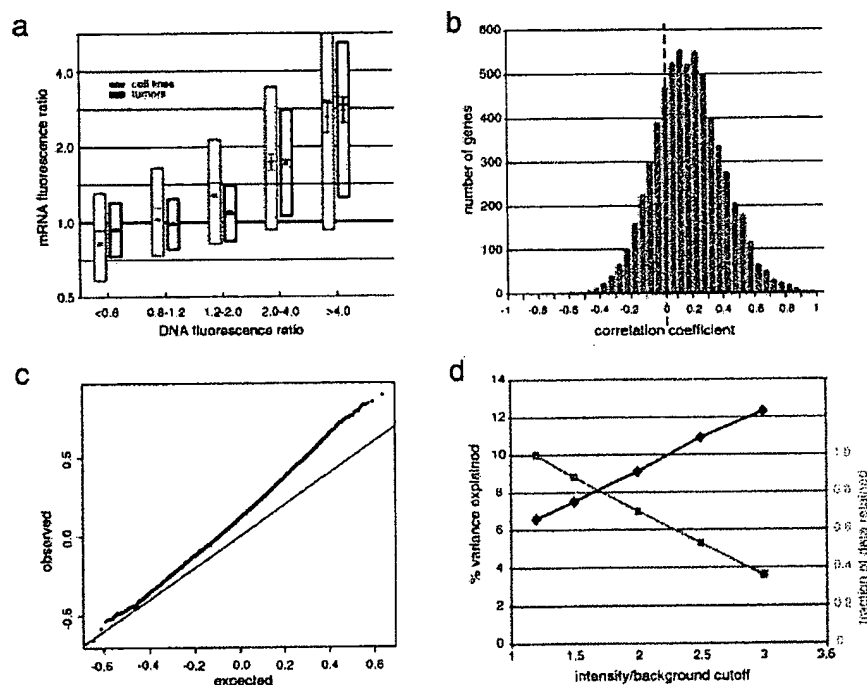
Fig. 4. Genome-wide influence of DNA copy number alterations on mRNA levels. (a) For breast cancer cell lines (gray) and tumor samples (black), both mean-centered mRNA fluorescence ratio ($\log_2$ scale) quartiles (box plots indicate 25th, 50th, and 75th percentile) and averages (diamonds; Y-value error bars indicate standard errors of the mean) are plotted for each of five classes of genes, representing DNA deletion (tumor/normal ratio < 0.8), no change (0.8–1.2), low- (1.2–2), medium- (2–4), and high-level (>4) amplification. P values for pair-wise Student's t tests, comparing averages between adjacent classes (moving left to right), are 4×10^{-49}, 1×10^{-49}, 5×10^{-5}, 1×10^{-2} (cell lines), and 1×10^{-43}, 1×10^{-214}, 5×10^{-41}, 1×10^{-4} (tumors). (b) Distribution of correlations between DNA copy number and mRNA levels, for 6,095 different human genes across 37 breast tumor samples. (c) Plot of observed versus expected correlation coefficients. The expected values were obtained by randomization of the sample labels in the DNA copy number data set. The line of unity is indicated. (d) Percent variance in gene expression (among tumors) directly explained by variation in gene copy number. Percent variance explained (black line) and fraction of data retained (gray line) are plotted for different fluorescence intensity/background (a rough surrogate for signal/noise) cutoff values. Fraction of data retained is relative to the 1.2 intensity/background cutoff. Details of the linear regression model used to estimate the fraction of variation in gene expression attributable to underlying DNA copy number alteration can be found in the supporting information (see *Estimating the Fraction of Variation in Gene Expression Attributable to Underlying DNA Copy Number Alteration*).

tumors that could be attributed to underlying variation in DNA copy number. From this analysis, we estimate that, overall, about 7% of all of the observed variation in mRNA levels can be explained directly by variation in copy number of the altered genes (Fig. 4d). We can reduce the effects of experimental measurement error on this estimate by using only that fraction of the data most reliably measured (fluorescence intensity/background >3); using that data, our estimate of the percent variation in mRNA levels directly attributed to variation in gene copy number increases to 12% (Fig. 4d). This still undoubtedly represents a significant underestimate, as the observed variation in global gene expression is affected not only by true variation in the expression programs of the tumor cells themselves, but also by the variable presence of non-tumor cell types within clinical samples.

Discussion

This genome-wide, array CGH analysis of DNA copy number alteration in a series of human breast tumors demonstrates the usefulness of defining amplicon boundaries at high resolution (gene-by-gene), and quantitatively measuring amplicon shape, to assist in locating and identifying candidate oncogenes. By analyzing mRNA levels in parallel, we have also discovered that changes in DNA copy number have a large, pervasive, direct effect on global gene expression patterns in both breast cancer

cell lines and tumors. Although the DNA microarrays used in our analysis may display a bias toward characterized and/or highly expressed genes, because we are examining such a large fraction of the genome (approximately 20% of all human genes), and because, as detailed above, we are likely underestimating the contribution of DNA copy number changes to altered gene expression, we believe our findings are likely to be generalizable (but would nevertheless still be remarkable if only applicable to this set of ~6,100 genes).

In budding yeast, aneuploidy has been shown to result in chromosome-wide gene expression biases (13). Two recent studies have begun to examine the global relationship between DNA copy number and gene expression in cancer cells. In agreement with our findings, Phillips *et al.* (14) have shown that with the acquisition of tumorigenicity in an immortalized prostate epithelial cell line, new chromosomal gains and losses resulted in a statistically significant respective increase and decrease in the average expression level of involved genes. In contrast, Platzer *et al.* (15) recently reported that in metastatic colon tumors only ~4% of genes within amplified regions were found more highly (>2-fold) expressed, when compared with normal colonic epithelium. This report differs substantially from our finding that 62% of highly amplified genes in breast cancer exhibit at least 2-fold increased expression. These contrasting findings may reflect methodological differences between the

studies. For example, the study of Platzer et al. (15) may have systematically under-measured gene expression changes. In this regard it is remarkable that only 14 transcripts of many thousand residing within unamplified chromosomal regions were found to exhibit at least 4-fold altered expression in metastatic colon cancer. Additionally, their reliance on lower-resolution chromosomal CGH may have resulted in poorly delimiting the boundaries of high-complexity amplicons, effectively overcalling regions with amplification. Alternatively, the contrasting findings for amplified genes may represent real biological differences between breast and metastatic colon tumors; resolution of this issue will require further studies.

Our finding that widespread DNA copy number alteration has a large, pervasive and direct effect on global gene expression patterns in breast cancer has several important implications. First, this finding supports a high degree of copy number-dependent gene expression in tumors. Second, it suggests that most genes are not subject to specific autoregulation or dosage compensation. Third, this finding cautions that elevated expression of an amplified gene cannot alone be considered strong independent evidence of a candidate oncogene's role in tumorigenesis. In our study, fully 62% of highly amplified genes demonstrated moderately or highly elevated expression. This highlights the importance of high-resolution mapping of amplicon boundaries and shape [to identify the "driving" gene(s) within amplicons (16)], on a large number of samples, in addition to functional studies. Fourth, this finding suggests that analyzing the genomic distribution of expressed genes, even within existing microarray gene expression data sets, may permit the inference of DNA copy number aberration, particularly aneuploidy (where gene expression can be averaged across large chromosomal regions; see Fig. 3 and supporting information). Fifth, this finding implies that a substantial portion of the phenotypic uniqueness (and by extension, the heterogeneity in clinical behavior) among patients' tumors may be traceable to underlying variation in DNA copy number. Sixth, this finding supports a possible role for widespread DNA copy number alteration in tumorigenesis (17, 18), beyond the amplification of specific oncogenes and deletion of specific tumor suppressor genes. Widespread DNA copy number alteration, and the concomitant widespread imbalance in gene expression, might disrupt critical stochiometric relationships in cell metabolism and physiology (e.g., proteosome, mitotic spindle), possibly promoting further chromosomal instability and directly contributing to tumor development or progression. Finally, our findings suggest the possibility of cancer therapies that exploit specific or global imbalances in gene expression in cancer.

1. Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F. & Pinkel, D. (1992) Science 258, 818–821.
2. Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W. & Waldman, F. M. (1994) Proc. Natl. Acad. Sci. USA 91, 2156–2160.
3. Tirkkonen, M., Tanner, M., Karhu, R., Kallioniemi, A., Isola, J. & Kallioniemi, O. P. (1998) Genes Chromosomes Cancer 21, 177–184.
4. Forozan, F., Mahlamaki, E. H., Monni, O., Chen, Y., Veldman, R., Jiang, Y., Gooden, G. C., Ethier, S. P., Kallioniemi, A. & Kallioniemi, O. P. (2000) Cancer Res. 60, 4519–4525.
5. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T. & Lichter, P. (1997) Genes Chromosomes Cancer 20, 399–407.
6. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., et al. (1998) Nat. Genet. 20, 207–211.
7. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. & Brown, P. O. (1999) Nat. Genet. 23, 41–46.
8. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. (2000) Nature (London) 406, 747–752.
9. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., et al. (2001) Proc. Natl. Acad. Sci. USA 98, 10869–10874.
10. Schuler, G. D. (1997) J. Mol. Med. 75, 694–698.
11. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) Nature (London) 409, 860–921.
12. Fejzo, M. S., Godfrey, T., Chen, C., Waldman, F. & Gray, J. W. (1998) Genes Chromosomes Cancer 22, 105–113.
13. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., et al. (2000) Nat. Genet. 25, 333–337.
14. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., et al. (2001) Cancer Res. 61, 8143–8149.
15. Platzer, P., Upender, M. B., Wilson, K., Willis, J., Lutterbaugh, J., Nosrati, A., Willson, J. K., Mack, D., Ried, T. & Markowitz, S. (2002) Cancer Res. 62, 1134–1138.
16. Albertson, D. G., Ylstra, B., Segraves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W. L., Gray, J. W. & Pinkel, D. (2000) Nat. Genet. 25, 144–146.
17. Li, R., Yerganian, G., Duesberg, P., Kraemer, A., Willer, A., Rausch, C. & Hehlmann, R. (1997) Proc. Natl. Acad. Sci. USA 94, 14506–14511.
18. Rasnick, D. & Duesberg, P. H. (1999) Biochem. J. 340, 621–630.

Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer[1,2]

Elizabeth Hyman,[3] Päivikki Kauraniemi,[3] Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Outi Monni, Abdel Elkahloun, Olli-P. Kallioniemi, and Anne Kallioniemi[4]

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E. H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E. H., P. K., S. H., M. W., S. M., E. R., M. R., A. E., O. K., A. K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33520 Tampere, Finland [P. K., A. K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S. H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G. S.]; and Biomedicum Biochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O. M.]

ABSTRACT

Genetic changes underlie tumor progression and may lead to cancer-specific expression of critical genes. Over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the *HOXB7* gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.2% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1–6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as *ERBB2* and *EGFR* (7, 8), in breast cancer and other solid tumors. Besides amplifications of known oncogenes, over
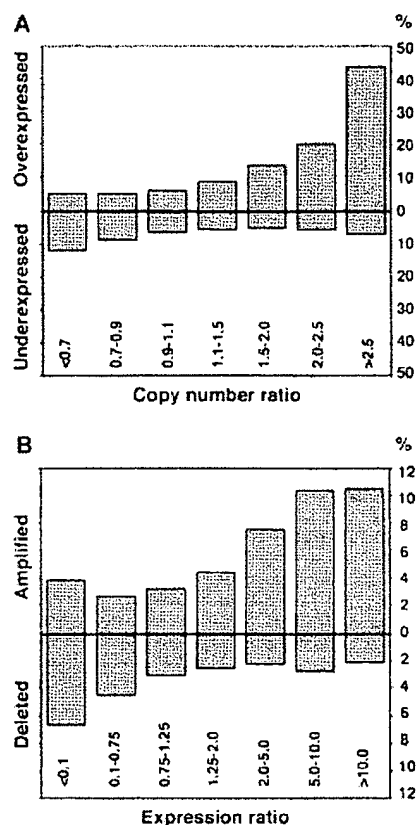
Fig. 1. Impact of gene copy number on global gene expression levels. *A*, percentage of over- and underexpressed genes (*Y axis*) according to copy number ratios (*X axis*). Threshold values used for over- and underexpression were >2.184 (global upper 7% of the cDNA ratios) and <0.4826 (global lower 7% of the expression ratios). *B*, percentage of amplified and deleted genes according to expression ratios. Threshold values for amplification and deletion were >1.5 and <0.7.

20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH[5] (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (*a*) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (*b*) identify and characterize those genes whose mRNA expres-

sion is most significantly associated with amplification of the corresponding genomic template.

MATERIALS AND METHODS

Breast Cancer Cell Lines. Fourteen breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

Copy Number and Expression Analyses by cDNA Microarrays. The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11–13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20 μg of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14–18 h with *Alu*I and *Rsa*I (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six μg of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty μg of reference RNA were labeled with Cy3-dUTP and 3.5 μg of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (*i.e.*, copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units)

were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

Statistical Analysis of CGH and cDNA Microarray Data. To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight, w_g, for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where m_{g1}, σ_{g1} and m_{g0}, σ_{g0} denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by α. A low α (<0.05) indicates a strong association between gene expression and amplification.

Genomic Localization of cDNA Clones and Amplicon Mapping. Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.[6] A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz's GoldenPath database.[7] The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

---

[6] Internet address: http://research.nhgri.nih.gov/microarray/downloadable_cdna.html.
[7] Internet address: www.genome.ucsc.edu.

Table 1 *Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray*

Location	Start (Mb)	End (Mb)	Size (Mb)
1p13	132.79	132.94	0.2
1q21	173.92	177.25	3.3
1q22	179.28	179.57	0.3
3p14	71.94	74.66	2.7
7p12.1-7p11.2	55.62	60.95	5.3
7q31	125.73	130.96	5.2
7q32	140.01	140.68	0.7
8q21.11-8q21.13	86.45	92.46	6.0
8q21.3	98.45	103.05	4.6
8q23.3-8q24.14	129.88	142.15	12.3
8q24.22	151.21	152.16	1.0
9p13	38.65	39.25	0.6
13q22-q31	77.15	81.38	4.2
16q22	86.70	87.62	0.9
17q11	29.30	30.85	1.6
17q12-q21.2	39.79	42.80	3.0
17q21.32-q21.33	52.47	55.80	3.3
17q22-q23.3	63.81	69.70	5.9
17q23.3-q24.3	69.93	74.99	5.1
19q13	40.63	41.40	0.8
20q11.22	34.59	35.85	1.3
20q13.12	44.00	45.62	1.6
20q13.12-q13.13	46.45	49.43	3.0
20q13.2-q13.32	51.32	59.12	7.8

CGH were validated, with 1q21, 17q12-q21.2, 17q22-q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65-39.25 Mb), and 17q21.3 (52.47-55.80 Mb).

Direct Identification of Putative Amplification Target Genes. The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2C shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22-q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates *EGFR* as the most highly overexpressed and amplified gene at 7p11-p12 (Fig. 3A). In BT-474, the two known amplicons at 17q12 and 17q22-q23 contained numerous highly overexpressed genes (Fig. 3B). In addition, several genes, including the homeobox genes *HOXB2* and *HOXB7*, were highly amplified in a previously undescribed independent amplicon at 17q21.3. *HOXB7* was systematically amplified (as validated by FISH, Fig. 3B, inset) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

extended to include neighboring nonamplified clones (ratio, <1.5). The amplicon size determination was partially dependent on local clone density.

FISH. Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and Spectrum-Orange-labeled probe for *EGFR* was obtained from Vysis. SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

RT-PCR. The *HOXB7* expression level was determined relative to *GAPDH*. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. *HOXB7* primers were 5'-GAGCAGAGGGACTCGGACTT-3' and 5'-GCGTCAGGTAGCGATTGTAG-3'.

RESULTS

Global Effect of Copy Number on Gene Expression. 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (i.e., belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1A). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1B). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

Identification of Distinct Breast Cancer Amplicons. Base-pair locations obtained for 11,994 cDNAs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. A). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal

a MDA-468 Chr. 7 — EGFR, SEC61G (Copy number ratio vs Basepairs ×10⁸)

b BT-474 Chr. 17 — MLN64, ERBB2, HOXB7, PPM1D, CLTC, HOXB2, MUL, RPS6KB1 (Copy number ratio vs Basepairs ×10⁸)

Fig. 3. Annotation of gene expression data on CGH microarray profiles. *A*, genes in the 7p11-p12 amplicon in the MDA-468 cell line are highly expressed (red dots) and include the *EGFR* oncogene. *B*, several genes in the 17q12, 17q21.3, and 17q23 amplicons in the BT-474 breast cancer cell line are highly overexpressed (red) and include the *HOXB7* gene. The data labels and color coding are as indicated for Fig. 2C. Insets show chromosomal CGH profiles for the corresponding chromosomes and validation of the increased copy number by interphase FISH using *EGFR* (red) and chromosome 7 centromere probe (green) to MDA-468 (A) and *HOXB7*-specific probe (red) and chromosome 17 centromere (green) to BT-474 cells (B).

Copy number ratio   Expression ratio

Gene name	Locus	Alpha
adaptor-related protein complex 4, beta 1 subunit	1p13	0.014
tripartite motif-containing 33	1p13	0.014
breast carcinoma amplified sequence 2	1p13	0.011
NRAS-related gene	1p13	0.010
renin	1q32	0.046
KIAA0456 protein	1q32	0.022
rab3 GTPase-activating protein	1q41	0.011
eukaryotic translation initiation factor 3, subunit 5	2p16	0.010
replication factor C	3q27	0.005
retinoic acid induced 14	5p13	0.010
mitogen-activated protein kinase kinase kinase 1	5q11	0.023
epidermal growth factor receptor	7p11	0.011
glioblastoma amplified sequence	7p11	0.012
CGI-62 protein	8q21	0.014
zinc finger protein RINZF	8q21	0.019
syndecan 2	8q22	0.011
eukaryotic translation initiation factor 3, subunit 6	8q23	0.039
v-myc	8q24	0.020
stomatin-like 2	9p13	0.010
protein phosphatase 1	11q13	0.046
NADH dehydrogenase	11q13	0.008
alanyl-tRNA synthetase	16q22	0.006
breast cancer anti-estrogen resistance 1	16q23	0.022
zinc finger protein 144 (Mel-18)	17q12	0.002
LIM and SH3 protein 1	17q12	0.012
hypothetical protein FLJ20840	17q21	0.007
steroidogenic acute regulatory protein related	17q21	0.001
phenylethanolamine N-methyltransferase	17q21	0.003
v-erb-b2	17q21	0.003
MLN51 protein	17q21	0.011
homeo box B2	17q21	0.030
homeo box B7	17q21	0.021
RAD51 (S. cerevisiae) homolog C	17q22	0.015
Multibrey nanism	17q22	0.003
clathrin, heavy polypeptide (Hc)	17q23	0.002
ribosomal protein S6 kinase	17q23	0.003
amyloid beta precursor protein -binding protein 2	17q23	0.001
protein phosphatase 1D magnesium-dependent	17q23	0.001
angiopoietin-like 4	19p13	0.022
nuclear receptor coactivator RAP250	20q11	0.001
acetyl-CoA synthetase	20q11	0.028
topoisomerase (DNA) I	20q12	0.011
MYBL2	20q13	0.010
serine/threonine kinase 4	20q13	0.021
nuclear receptor coactivator 3	20q13	0.010
prefoldin 4	20q13	0.019
transcription factor AP-2 gamma	20q13	0.003
bone morphogenetic protein 7	20q13	0.001
zinc finger protein 276	22q12	0.002
KIAA0443 gene product	Xq22	0.007

Copy number ratios
- top 2%
- top 3-5%
- middle 90%
- bottom 3-5%
- bottom 2%

Expression ratios
- top 2%
- top 3-7%
- middle 86%
- bottom 3-7%
- bottom 2%

Fig. 4. List of 50 genes with a statistically significant correlation (α value <0.05) between gene copy number and gene expression. Name, chromosomal location, and the α value for each gene are indicated. The genes have been ordered according to their position in the genome. The color maps on the *right* illustrate the copy number and expression ratio patterns in the 14 cell lines. The key to the color code is shown at the *bottom* of the graph. *Gray squares*, missing values. The complete list of 270 genes is shown in supplemental Fig. B.

amplification was validated to be present in 10.2% of 363 primary breast cancers by FISH to a tissue microarray and was associated with poor prognosis of the patients ($P = 0.001$).

Statistical Identification and Characterization of 270 Highly Expressed Genes in Amplicons. Statistical comparison of expression levels of all genes as a function of gene amplification identified 270 genes whose expression was significantly influenced by copy number across all 14 cell lines (Fig. 4, Supplemental Fig. B). According to the gene ontology data,[8] 91 of the 270 genes represented hypothetical proteins or genes with no functional annotation, whereas 179 had associated functional information available. Of these, 151 (84%) are implicated in apoptosis, cell proliferation, signal transduction, and transcription, whereas 28 (16%) had functional annotations that could not be directly linked with cancer.

DISCUSSION

The importance of recurrent gene and chromosome copy number changes in the development and progression of solid tumors has been characterized in >1000 publications applying CGH[9] (9, 10), as well as in a large number of other molecular cytogenetic, cytogenetic, and molecular genetic studies. The effects of these somatic genetic changes on gene expression levels have remained largely unknown, although a few studies have explored gene expression changes occurring in specific amplicons (15, 19–21). Here, we applied genome-wide cDNA microarrays to identify transcripts whose expression changes were attributable to underlying gene copy number alterations in breast cancer.

The overall impact of copy number on gene expression patterns was substantial with the most dramatic effects seen in the case of high-

---

level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22–24).

The CGH microarray analysis identified 24 independent breast cancer amplicons. We defined the precise boundaries for many amplicons detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1–2 Mb) or close proximity to other larger amplicons. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the over-expression of the *HOXB7* and *HOXB2* genes. The homeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). *HOXB7* transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29–32). The present results imply that gene amplification may be a prominent mechanism for overexpressing *HOXB7* in breast cancer and suggest that *HOXB7* contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of *HOXB7* in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as *HER-2*, *MYC*, *EGFR*, ribosomal protein s6 kinase, and *AIB3*, but also numerous novel genes such as *NRAS-related gene* (1p13), *syndecan-2* (8q22), and *bone morphogenic protein* (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenetic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly once every 267 kb. This analysis provided: (*a*) evidence of a prominent global influence of copy number changes on gene expression levels; (*b*) a high-resolution map of 24 independent amplicons in breast cancer; and (*c*) identification of a set of 270 genes, the overexpression of which was statistically attributable to gene amplification. Characterization of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer, including a clinical association

between *HOXB7* amplification and poor patient prognosis. Overall, our results illustrate how the identification of genes activated by gene amplification provides a powerful approach to highlight genes with an important role in cancer as well as to prioritize and validate putative targets for therapy development.

REFERENCES

1. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeck, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (Wash. DC), *286:* 531–537, 1999.
2. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature (Lond.), *403:* 503–511, 2000.
3. Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., *et al.* Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature (Lond.), *406:* 536–540, 2000.
4. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* Molecular portraits of human breast tumours. Nature (Lond.), *406:* 747–752, 2000.
5. Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. Nature (Lond.), *412:* 822–826, 2001.
6. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA, *98:* 10869–10874, 2001.
7. Ross, J. S., and Fletcher, J. A. The *HER-2/neu* oncogene: prognostic factor, predictive factor and target for therapy. Semin. Cancer Biol., *9:* 125–138, 1999.
8. Arteaga, C. L. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. J. Clin. Oncol., *19:* 32–40, 2001.
9. Knuutila, S., Bjorkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., El-Rifai, W., *et al.* DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. Am. J. Pathol., *152:* 1107–1123, 1998.
10. Knuutila S., Autio K., and Aalto Y. Online access to CGH data of DNA sequence copy number changes. Am. J. Pathol., *157:* 689, 2000.
11. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat. Genet., *14:* 457–460, 1996.
12. Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res., *6:* 639–645, 1996.
13. Mousses, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Baxevanis, A., Meltzer, P. S., and Trent, J. M. Gene expression analysis by cDNA microarrays. In: F. J. Livesey and S. P. Hunt (eds.), Functional Genomics, pp. 113–137. Oxford: Oxford University Press, 2000.
14. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat. Genet., *23:* 41–46, 1999.
15. Monni, O., Bärlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. Proc. Natl. Acad. Sci. USA, *98:* 5711–5716, 2001.
16. Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. J. Biomed. Optics, *2:* 364–374, 1997.
17. Bärlund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Torhorst, J., Haas, P., Bucher, C., Sauter, G., *et al.* Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. J. Natl. Cancer Inst., *92:* 1252–1259, 2000.
18. Andersen, C. L., Hostetter, G., Grigoryan, A., Sauter, G., and Kallioniemi, A. Improved procedure for fluorescence *in situ* hybridization on tissue microarrays. Cytometry, *45:* 83–86, 2001.
19. Kauraniemi, P., Bärlund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. Cancer Res., *61:* 8235–8240, 2001.
20. Clark, J., Edwards, S., John, M., Flohr, P., Gordon, T., Maillard, K., Giddings, I., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., and Cooper, C. S. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. Genes Chromosomes Cancer, *34:* 104–114, 2002.
21. Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kokkola, A., Moskaluk, C., Frierson, H., Powell, S. M., Knuutila, S., Kallioniemi, A., and El-Rifai, W. Targets of gene amplification and overexpression at 17q in gastric cancer. Cancer Res., *62:* 2625–2629, 2002.
22. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H., and Marton M. J.

Widespread aneuploidy revealed by DNA microarray expression profiling. Nat. Genet., *25:* 333–337, 2000.

23. Virtaneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de La Chapelle, A., and Krahe, R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc. Natl. Acad. Sci. USA, *98:* 1124–1129, 2001.

24. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., Linehan, W. M., Cunha, G. R., and Ried, T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. Cancer Res., *61:* 8143–8149, 2001.

25. Bärlund, M., Tirkkonen, M., Forozan, F., Tanner, M. M., Kallioniemi, O. P., and Kallioniemi, A. Increased copy number at 17q22–q24 by CGH in breast cancer is due to high-level amplification of two separate regions. Genes Chromosomes Cancer, *20:* 372–376, 1997.

26. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjärvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi O. P. Independent amplification and frequent co-amplification of three nonsyntenic regions

27. Cillo, C., Faiella, A., Cantile, M., and Boncinelli, E. Homeobox genes and cancer. Exp. Cell Res., *248:* 1–9, 1999.

28. Cillo, C., Cantile, M., Faiella, A., and Boncinelli, E. Homeobox genes in normal and malignant cells. J. Cell. Physiol., *188:* 161–169, 2001.

29. Care, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. Mol. Cell. Biol., *16:* 4842–4851, 1996.

30. Care, A., Silvani, A., Meccia, E., Mattia, G., Peschle, C., and Colombo, M. P. Transduction of the SkBr3 breast carcinoma cell line with the *HOXB7* gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. Oncogene, *16:* 3285–3289, 1998.

31. Care, A., Felicetti, F., Meccia, E., Bottero, L., Parenza, M., Stoppacciaro, A., Peschle, C., and Colombo, M. P. HOXB7: a key factor for tumor-associated angiogenic switch. Cancer Res., *61:* 6532–6539, 2001.

32. Naora, H., Yang, Y. Q., Montz, F. J., Seidman, J. D., Kurman, R. J., and Roden, R. B. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. Proc. Natl. Acad. Sci. USA, *98:* 4060–4065, 2001.

on the long arm of chromosome 20 in human breast cancer. Cancer Res., *56:* 3441–3445, 1996.

Targets of Gene Amplification and Overexpression at 17q in Gastric Cancer[1]

Asta Varis,[2] Maija Wolf,[2] Outi Monni,[2] Marja-Leena Vakkari, Arto Kokkola, Chris Moskaluk, Henry Frierson, Jr., Steven M. Powell, Sakari Knuutila, Anne Kallioniemi, and Wa'el El-Rifai[3]

Department of Medical Genetics [A. V., M-L. V., S. K.], Biomedicum Biochip Center [O. M.], and Department of Surgery [A. K.], University of Helsinki and Helsinki University Central Hospital, Helsinki 00290, Finland; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [M. W., O. M., A. K.]; Laboratory of Cancer Genetics, University of Tampere, Tampere 33014, Finland [A. K.]; and Department of Pathology [C. M., H. F.] and Digestive Health Center of Excellence [S. M. P., W. E-R.], University of Virginia Health System, Charlottesville, Virginia 22908-0708

ABSTRACT

DNA copy number gains and amplifications at 17q are frequent in gastric cancer, yet systematic analyses of the 17q amplicon have not been performed. In this study, we carried out a comprehensive analysis of copy number and expression levels of 636 chromosome 17-specific genes in gastric cancer by using a custom-made chromosome 17-specific cDNA microarray. Analysis of DNA copy number changes by comparative genomic hybridization on cDNA microarray revealed increased copy numbers of 11 known genes (*ERBB2, TOP2A, GRB7, ACLY, PIP5K2B, MPRL45, MKP-L, LHX1, MLN51, MLN64,* and *RPL27*) and seven expressed sequence tags (ESTs) that mapped to 17q12-q21 region. To investigate the genes transcribed at the 17q, we performed gene expression analyses on an identical cDNA microarray. Our expression analysis showed overexpression of 8 genes (*ERBB2, TOP2A, GRB2, AOC3, AP2B1, KRT14, JUP,* and *ITGA3*) and two ESTs. Of the commonly amplified transcripts, an uncharacterized EST AA552509 and the *TOP2A* gene were most frequently overexpressed in 82% of the samples. Additional studies will be initiated to understand the possible biological and clinical significance of these genes in gastric cancer development and progression.

INTRODUCTION

Gastric carcinoma is one of the most common malignancies worldwide and is the second most frequent cause of cancer-related death (1). Moreover, cardia, gastroesophageal junction, and esophageal adenocarcinomas have the most rapidly rising incidence of all visceral malignancies in the United States and Western world for reasons that are unclear (2). Previous studies have documented the importance of genetic alterations affecting known oncogenes, tumor suppressor genes, and mismatch repair genes in the development of gastric cancer (3, 4). Several genes, such as *cMET, ERBB2, MYC,* and *MDM2,* are amplified in 10–25% of tumors, and their amplification is associated with advanced disease (3, 5). Comprehensive DNA copy number analyses of gastric cancers using CGH[4] have demonstrated recurrent DNA copy number changes on several chromosomal regions. Gains at 17q have been shown to be frequent in gastric cancers (6). However, the critical regions of genetic alterations are large, and the target genes for amplification at 17q remain unknown.

Characterization of the chromosomal regions involved in DNA copy number changes is likely to reveal genes important for the development of gastric cancer. In the present study, we used a custom-made chromosome 17-specific cDNA microarray to systematically evaluate copy numbers and expression levels of genes at 17q in gastric carcinomas.

MATERIALS AND METHODS

Samples. Sixteen gastric cancer xenografts, four gastric cancer cell lines (CRL-5822, CRL-5974, CRL-5973, and CRL-1739) from the American Type Culture Collection (Manassas, VA), and five primary gastric cancers were used in this study. The cell line (CRL-1739) with normal DNA copy number of chromosome 17 was included as a control in Northern blot hybridizations. The cell lines were cultured under recommended conditions. Xenografting of gastric cancers was performed as described earlier (7). All tumors included in this study were dissected and verified histologically to be composed predominantly of neoplastic tissues. We have earlier characterized the DNA copy numbers of the cell lines and xenografts using "chromosomal" CGH. The details of the DNA copy numbers of the xenografts have been reported elsewhere (7). Fig. 1*A* summarizes the chromosomal CGH results for chromosome 17.

Chromosome 17-specific cDNA Microarray. The construction of the chromosome 17-specific cDNA microarray has been described previously (8). Briefly, the cDNA microarray contained a total of 636 clones, including 88 house keeping genes, 201 known genes from chromosome 17, and 435 EST clones from radiation hybrid map intervals D17S933-D17S930 (293–325 cR, the 17q12-q21 region) and D17S791-D17S795 (333–435 cR, the 17q23-q24 region). The preparation and printing of the cDNA clones on glass slides were performed as described elsewhere (9).

Copy Number and Expression Analyses by cDNA Microarrays. Genomic DNA was extracted from eight xenografts (X11, X27, X57, X71, X75, X79, X83, and X95) and three cell lines (CRL-5822, CRL-5973, and CRL-5974). All cases had gains or high-level amplification at 17q by chromosomal CGH (Fig. 1). Normal genomic DNA was used as a reference in all experiments. Copy number analysis using CGH microarray was performed as described previously (8, 10). Briefly, 20 μg of genomic DNA were digested for 14–18 h with *AluI* and *RsaI* restriction enzymes (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Digested gastric cancer test DNA (6 μg) was labeled with Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, NJ) and 6 μg of reference DNA with Cy5-dUTP using Bioprime Labeling kit (Life Technologies, Inc.). Hybridization was done according to the protocol by Pollack *et al.* (10) and posthybridization washes as described previously (11).

Total RNA was extracted from eight xenografts (X43, X49, X57, X68, X75, X76, X80, and X95) and three gastric cancer cell lines (CRL-5822, CRL-5973, and CRL-5974) by using RNeasy kit (Qiagen, GmbH, Hilden, Germany). A pool of four normal gastric epithelial tissue samples, enriched for the epithelial layer of the stomach through dissection and mucosal scrapping, was used as a standard reference in all experiments. Reference RNA (100 μg) was labeled with Cy5-dUTP and 80 μg of test RNA with Cy3-dUTP by use of oligodeoxythymidylate-primed polymerization by SuperScript II reverse transcriptase (Life Technologies, Inc.). The labeled cDNAs were hybridized on microarrays as described previously (11, 12).

For both the copy number and expression analyses, the fluorescence intensities at the cDNA targets were measured by using a laser confocal scanner (Agilent Technologies, Palo Alto, CA). The fluorescent images from the test and control hybridizations were scanned separately, and the data were analyzed using the DEARRAY software (13). After the subtraction of background intensities, the average intensities of each spot in the test hybridization were divided by the average intensity of the corresponding spot in the control hybridization. On the basis of our earlier reports (8, 14), clones that showed a

[4] The abbreviations used are: CGH, comparative genomic hybridization; EST, expressed sequence tag; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; RT-PCR, reverse transcription-PCR; AP, adapter protein.
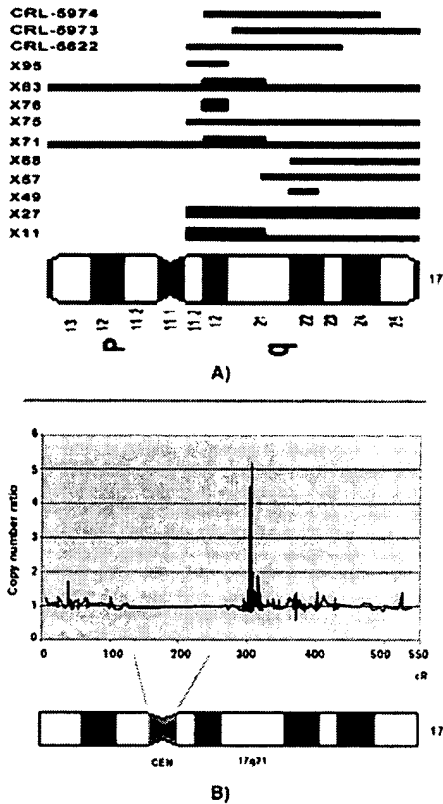
A)

B)

Fig. 1. DNA copy numbers in gastric cancer. *A,* summary of gains and high-level amplifications affecting chromosome 17 in gastric cancer xenografts and cell lines by chromosomal CGH. *Horizontal bars,* the extent of the copy number aberration in each sample. High-level amplifications are presented as *wide bars.* xenograft samples X43, X80, and cell line CRL-1739 had no detectable gains on chromosome 17. *B,* copy numbers survey of chromosome 17-specific genes in X83 gastric cancer xenograft by CGH microarray. The copy number ratios were plotted as a function of the position of the clones in the radiation hybrid map in cR scale. Individual data points were connected with a *line.* The chromosome 17 ideogram is shown *below* for visual comparison only.

copy number ratio ≥ 1.5 were considered as amplified, and clones that showed an expression ratio ≥ 3 were considered as overexpressed. Clones that showed such increased ratios in the self *versus* self control experiment were excluded from the analysis.

Northern Hybridization. Total RNA was extracted from four gastric cancer cell lines and two normal stomach specimens using the RNeasy kit (Qiagen, GmbH). The Northern hybridization was performed using standard methods. Briefly, 10 μg of total RNA were size-fractionated on a 1% agarose gel containing formaldehyde and transferred on a Nytran membrane (Schleicher & Schuel, Keene, NH). The membrane was prehybridized for 1 h at 65°C in Express hybridization solution (Clontech, Palo Alto, CA) together with sheared Herring sperm DNA (10 μg/ml; Research Genetics, Huntsville, AL). Sequence-verified cDNA inserts were labeled with P[32] by random priming (Prime-It; Stratagene, La Jolla, Ca). Hybridization was performed in the Express hybridization solution (Clontech) at 65°C overnight followed by washes in 2 × SCC/SDS solutions. Signals were detected by autoradiography. The normal gastric tissues and CRL-1739 cell line (normal chromosome 17 on CGH) were used as control samples. A GAPDH cDNA was used as a control probe.

Multiplex RT-PCR. Multiplex RT-PCR was used to validate the cDNA array results for the two most overexpressed genes (*EST AA552509* and *TOP2A*) using seven xenografted and six primary gastric cancer samples. For reference expression, a pool of normal gastric epithelial tissues obtained from different individuals was used. Primary tumors of four xenografts were included in the analyses. mRNA was purified from the tissues using mRNeasy (Qiagen), and cDNA synthesis was performed using Advantage RT-for-PCR Kit (Clontech). In each PCR reaction, primers for the human *GAPDH* gene were used as an internal reference. The PCR reactions were done using standard protocol for 28 cycles. We confirmed the reproducibility of the method by repeating the RT-PCR twice, and the results were consistent. The primers used for the RT-PCR were obtained from GeneLink (Hawthorne, NY), and their sequences are available on request.

RESULTS

Detailed Characterization of the 17q Amplification Using Chromosome-specific Microarray. Copy number levels of 636 chromosome 17-specific genes were evaluated by CGH microarray in eight xenografts (X11, X27, X57, X71, X75, X79, X83, and X95) and three gastric cancer cell lines (CRL-5822, CRL-5973, and CRL-5974) that

Table 1 *Summary of copy number ratios of 18 chromosome 17q12-q21 transcripts in eight xenografts and three cell lines of gastric cancer by CGH microarray[a]*

														CRL-5822	CRL-5973	CRL-5974
Gene	Unigene Id	Accession	Alignment[b]	Locus[b]	X11	X27	X57	X71	X75	X79	X83	X95			
MRPL45 Mitochondrial ribosomal protein L45	Hs. 19347	AI277785	38274220/51922787	17q12/17q21.3	1.4	1.5	1.1	1	2.4	1.3	1.6	1.6	2	1.4	1.2
MKP-1 like protein tyrosine phosphatase (MKP-L)	Hs. 91448	AA129677	38747279	17q12	1.9	1.6	0.8	1.4	1.1	1.1	1.2	2.8	1.4	1.4	1.7
LIM homeobox protein 1 (LHX1)	Hs. 157449	AI375565	39307916	17q12	1	1.3	2.1	1.1	2.8	1.2	1	2.8	1.4	2.5	2.1
Phosphatidylinositol-4-phosphate 5-kinase, type II, β (PIP5K2B)	Hs. 6335	H80263	40617731	17q12	1	0.9	1.6	1.7	1.3	1.3	0.9	1.7	1.4	2.8	0.8
EST	Hs. 91668	H16094	41911205	17q21.1	1.2	0.8	1.2	1	1.4	3.8	4.7	0.9	3.1	1.2	0.8
EST (FLJ20940 hypothetical protein)	Hs. 286192	AA552509	41868584	17q21.1	1.4	1.1	1.7	1.2	1.4	5.8	7.3	1.6	1.5	1.4	1.1
H. sapiens MLN64 mRNA	Hs. 77628	AA504615	41877246	17q21.1	1.1	1.1	1.2	1.2	1.2	5.5	4.2	1.1	3.1	1.4	1.1
V-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (ERBB2)	Hs. 323910	AA446928	41940229	17q21.1	1	1	1	1	1.1	2.6	2.1	1	1.7	1.1	1
EST	Hs. 46645	AA283905	41972680	17q21.1	1.3	1.1	1	1.3	0.9	5.1	10.4	1.3	2	0.8	1.3
EST	Hs. 318893	AA455291	41978415	17q21.1	1.1	0.9	1.1	1	0.9	7.8	2.5	0.8	3.2	1.4	0.9
Growth factor receptor-bound protein 7 (GRB7)	Hs. 86859	H53703	41989210	17q21.1	1.4	1	1	1.1	1	5.7	8.9	0.9	2.7	1.4	1
H. sapiens MLN51 mRNA	Hs. 83422	R52974	42331857	17q21.1	1.6	1.2	1.1	1.6	0.9	1.9	1.9	1.5	1.4	2.1	1.1
Topoisomerase (DNA) II α (170kD) (TOP2A)	Hs. 156346	AA026682	42521254	17q21.2	1.3	1.1	1.2	1.6	1	1.7	1.6	1.4	1.6	1.4	1.1
EST	Hs. 13268	AA514361	44056922	17q21.2	1.2	1.1	1	1.9	1.3	2.3	1.2	1.8	1.5	1.8	1.5
ATP citrate lyase (ACLY)	Hs. 174140	R55974	44075311	17q21.2	1.2	1	1	1.6	1	1.6	1.1	1.6	1.4	1.6	2.3
EST	Hs. 38039	H62271	44574937	17q21.2	5.7	2.9	2.3	6.7	3.1	1	1.7	2.5	3	4.3	2.7
Ribosomal protein L27 (RPL27)	Hs. 111611	AA190881	45301136	17q21.2	4.6	2.5	2.2	6.1	1.7	0.9	2	2	2.1	1.4	2.9
EST (DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 8, DDX8)	Hs. 171872	AI540663	46054957	17q21.3	1.4	1.7	0.5	1.4	0.9	2.9	1.6	1.2	1.9	3.7	1.1

[a] Copy number ratios above the 1.5 threshold are shown in bold.

[b] Alignment (bp position) and locus are shown according to Santa Cruz August freeze 2001 assembly.

Table 2 *Summary of expression levels of 10 chromosome 17q12-q21 transcripts in eight xenografts and three cell lines of gastric cancer by cDNA microarray[a]*

| Gene | Unigene Id | Accession | Alignment[b] | Locus[b] | Samples | | | | | | | | | | |
					X43	X49	X57	X68	X75	X76	X80	X95	CRL-5822	CRL-5973	CRL-5974
Adaptor-related protein complex 2, β1 subunit (AP2B1)	Hs. 74626	H29927	37327700	17q12	1.1	1	2.9	3.1	0.9	1.1	0.9	0.8	**4.8**	**7.4**	**4.6**
EST (Hypothetical protein FLJ20940)	Hs. 286192	AA552509	41868584	17q21.1	**21.9**	**4.5**	**6.4**	**7.6**	**10.2**	**10**	**17.3**	0.6	**12.1**	**11.2**	0.6
V-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (ERBB2)	Hs. 323910	AA446928	41940229	17q21.1	1	1	**3**	**3.7**	1.4	0.9	1.3	0.7	**24.6**	0.7	1
Topoisomerase (DNA) II α (170kD) (TOP2A)	Hs. 156346	AA026682	42521254	17q21.2	**4.1**	**6.1**	**16**	**4.5**	2.8	**6.6**	**3**	1.4	**5.6**	**7.6**	**6.8**
Keratin 14 (KRT14)	Hs. 117729	H44127	43757143	17q21.2	**3.9**	1.4	1.1	1.6	**3.5**	1	1.2	0.6	**3.8**	1.3	0.8
Junction plakoglobin (JUP)	Hs. 2340	R06417	43994962	17q21.2	**3.1**	2.8	0.9	1.2	**4.3**	0.9	2.6	**3.4**	**5**	1.9	2.5
Amine oxidase, copper containing 3 (AOC3)	Hs. 198241	T77398	45078066	17q21.2	**4.6**	1.9	**4.5**	2.6	2.1	2.6	**4.2**	1.6	**3.1**	1.8	**5**
Integrin, α-3 (ITGA3)	Hs. 265829	AA424695	54688140	17q21.3	**4.8**	1.5	0.9	**4.5**	**3.4**	1.3	1.1	1.2	2.7	2.2	0.5
EST	Hs. 56105	AA284262	65817334	17q23.2	1	1.7	**5.2**	**3**	**15.6**	2.8	0.6	1.9	1.3	0.6	0.4
Growth factor receptor-bound protein 2 (GRB2)	Hs. 296381	AA449831	81840742	17q25.1	0.8	0.8	2.2	1.4	1.3	1.8	1.1	1.3	**3.1**	**5.6**	**3.5**

[a] Expression ratios above the 3 threshold are shown in bold.
[b] Alignment (bp position) and locus are shown according to Santa Cruz August freeze 2001 assembly.

showed gain or high-level amplification affecting chromosome 17 by chromosomal CGH (Fig. 1). CGH microarray analysis revealed increased DNA copy numbers (ratio ≥ 1.5) in three or more cases for 11 genes and seven ESTs that map to 17q12 (4 clones) and 17q21 (14 clones; Table 1). The amplified genes/ESTs were localized at 302–321 cR in the radiation hybrid map[5] (Fig. 1B) and between 38274220–46054957 bp at 17q, according to the University of California Santa Cruz's August freeze 2001 assembly of the human genome sequence.[6] The two most consistently amplified clones were EST (H62271) and ribosomal protein L27 (82%). Other frequently amplified genes included *TOP2A, EST AA552509,* and *ERBB2.* The details of the copy numbers and location of these genes/ESTs are listed in Table 1.

Gene Expression Profiling of 17q Using cDNA Microarrays. Parallel expression survey in eight xenografts (X43, X49, X57, X68, X75, X76, X80, and X95) and the three cell lines identified 10 transcripts at 17q whose expression was elevated (ratio ≥ 3) in at least three specimens, as compared with the normal gastric epithelial cells (Table 2; Fig. 2). Three of the commonly amplified sequences (*TOP2A, ERBB2,* and EST AA552509) that map to 17q21 were also overexpressed frequently in our cDNA expression analyses. The two most consistently affected transcripts were EST AA552509 (82%) and the *TOP2A* (82%).

Other frequently overexpressed genes included *AOC3* (45%), *JUP* (36%), *ERBB2* (27%), *ITAG3* (27%), and *KRT14* (27%) at 17q21 region, as well as *AP2B1* at 17q12, EST AA284262 at 17q23, and *GRB2* at 17q25 (Table 2; Fig. 2).

Northern Blotting. Northern analysis was used as an independent expression assay to validate the cDNA microarray results. Because of the limited availability of RNA from the xenografted tumors, only cell lines were analyzed. Three genes, *EST AA552509, TOP2A,* and *ERBB2,* that showed overexpression in one or more cell lines by cDNA microarray were selected for analysis. Results from the Northern analysis confirmed the cDNA microarray data. *ERBB2* was highly overexpressed in CRL-5822 cell line, *TOP2A* in all three cell lines, and EST AA552509 in CRL-5822 and CRL-5973 (Fig. 3). These genes were not expressed in the normal gastric epithelial sample or the gastric cell line (CRL-1739) that had normal chromosome 17 DNA copy numbers by chromosomal CGH (Fig. 3).

Multiplex RT-PCR. Expression analyses with RT-PCR showed elevated expression of *TOP2A* and EST AA552509 in all tested tumor samples, whereas no expression was seen in the pool of normal gastric epithelial tissues (Fig. 3). The xenografts and their corresponding primaries showed similar levels of expression.

DISCUSSION

Studies by chromosomal CGH have indicated that 17q is amplified frequently in gastric cancer. Here we used a custom-made cDNA
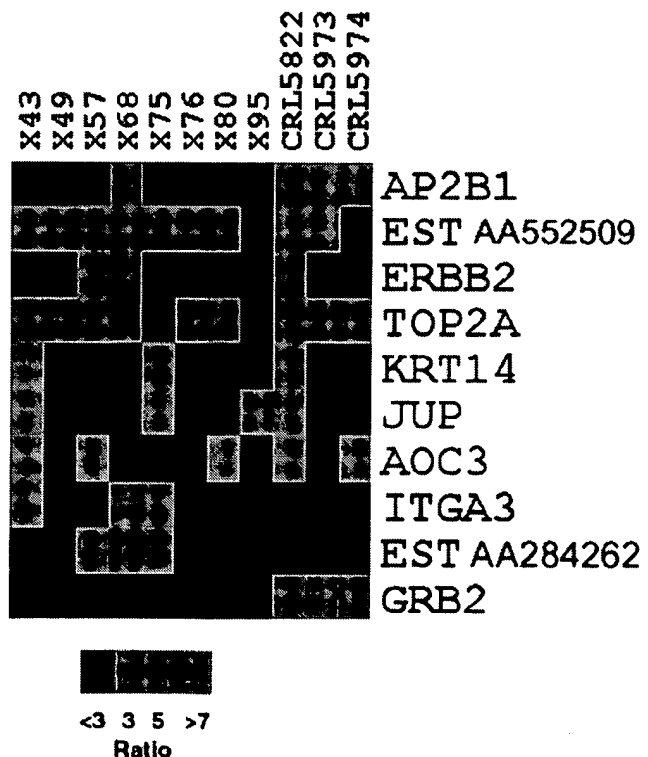
Fig. 2. Expression patterns of the most commonly overexpressed genes in gastric cancer xenografts and cell lines. Names of the genes are indicated on the *right.* Color coding for the expression ratios is shown *below* the graph. This image was created using Tree view software written by Michael Eisen, copyright 1998–1999, Stanford University.

[5] Internet address: http://www.ncbi.nlm.nih.gov/genemap.
[6] Internet address: http://genome.ucsc.edu.

Fig. 3. Validation of overexpressed genes in gastric cancer. *A*, Northern analysis of *TOP2A*, *ERBB2*, and EST AA552509 expression in normal gastric tissue and four gastric cancer cell lines. CRL-1739 had normal copy numbers by CGH. The size of each transcript is indicated on the *right* side of the corresponding picture. *GAPDH* was used as a loading control. *B*, expression analysis by multiplex RT-PCR in normal gastric tissue, seven xenografts (indicated by X-number), and six primary gastric cancers (indicated by G-number). Xenografts and their corresponding primary cancers have the same number. The names of the gene are shown on the *right*.
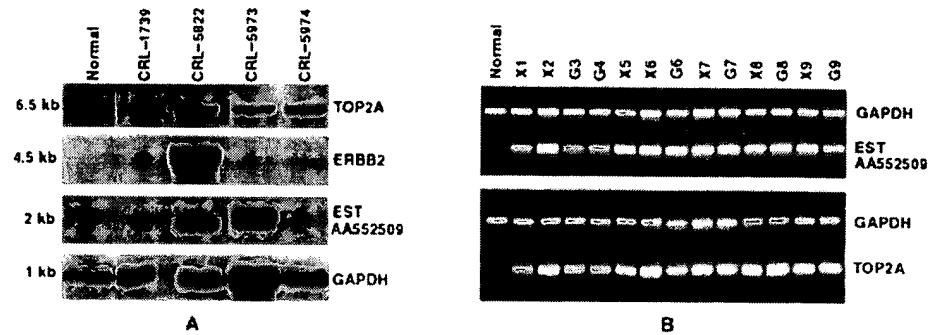
microarray that contained 636 cDNA clones from chromosome 17 to systematically analyze the copy number changes at 17q in eight gastric cancer xenografts and three cell lines. The CGH microarray analyses showed increased copy number ratios for 18 clones that were localized to the 17q12-q21 region. To identify those genes that are activated through increased copy number, we performed a comprehensive gene expression profiling using the same chromosome 17-specific cDNA microarray. Three of the commonly amplified transcripts (*TOP2A*, *ERBB2*, and EST AA552509) that map to 17q21 were overexpressed frequently in our analyses and might, therefore, represent putative amplification target genes in gastric cancer. The cDNA microarray results were validated using Northern and RT-PCR analyses.

The two most frequently overexpressed genes in our samples were the EST AA552509 and *TOP2A*. In addition, *ERBB2* was also amplified and overexpressed in >30% of tumors. Our data show that these genes are overexpressed in gastric cancers with no indication of their expression in normal gastric epithelial tissues. The overexpression of EST AA552509 has not been reported before and might be important for gastric carcinogenesis or have a possible value as a tumor marker or therapeutic target. On the other hand, the importance of *TOP2A*, and *ERBB2* in cancer, especially breast cancer, is well known (15, 16). *TOP2A* is an enzyme that catalyzes ATP-dependent strand-passing reactions and functions in DNA replication and chromosome condensation and segregation (17). *TOP2A* is a molecular target for many anticancer drugs (topo2 inhibitors). *ERBB2* is amplified frequently in breast cancer and has been shown to be an independent prognostic factor (18, 19). In breast cancer, *TOP2A* is often coamplified with *ERBB2* (20, 21). In our gastric adenocarcinomas, amplification and overexpression of *TOP2A* were independent of and also more frequent than *ERBB2*. Previous studies of *ERBB2* in gastric cancer have shown that the frequency of its overexpression varies from 9 to 38% (22, 23), which is in agreement with our findings. Our results provide additional evidence that clinical studies are required to determine the possibility that *TOP2A* and *ERBB2* are useful targets for cancer therapy in gastric cancer patients with these molecular alterations.

The up-regulation of *GRB2*, *JUP*, and *ITAG3* genes in the present study supports our earlier results that show these genes to be overexpressed in gastric cancer (7). Interestingly, studies in breast cancer suggest that *GRB2* may mediate transmission of *ERBB2* oncogenic signals, which in turn activate mitogen-activated protein kinase pathway (24, 25). *GRB2* is a widely expressed protein, which plays a crucial role in activation of several other growth factors (26).

KRT14, *AOC3*, and *AP2B1* were overexpressed in ≥3 of 11 of our gastric cancers. Copper-containing amino oxidases, such as *AOC3*, are involved in the catabolism of putrescine and histamine and are also involved in the regulation of growth and apoptosis (27). The AP2B1 is a member of AP complexes that function as vesicle coat components in different membrane traffic pathways. AP-2 complex associates with the plasma membrane and directs the internalization of

trafficking cell surface protein (28). However, there is no information about the possible role of these genes in cancer.

Our study has identified genes that are coamplified at 17q12 and 17q21 amplicons that are not altered transcriptionally in comparison of tumors to normal reference samples. The lack of correlation between some amplified genes and their expression profile suggests that these genes are not critical targets at the 17q amplicon but might be coamplified together with critical genes within the amplicon structure. We also found genes that were overexpressed but not amplified by CGH microarrays. These results in CGH microarray may be attributed to the resolution of CGH-based technologies. On the other hand, upstream gene regulation and/or mutations are known as important biological mechanisms in transcriptional regulation irrespective of gene copy number.

Comparison of this gastric cancer study with our earlier data from breast cancer using the same cDNA microarray revealed a different pattern of alterations affecting chromosome 17 (8, 14). In breast cancer, two common regions of increased copy number and expression, 17q12-q21 and 17q23, were observed. In addition, the genes influenced by the 17q12-q21 amplification in gastric cancer differed from those in breast cancer where *ERBB2* was among the most strongly affected (8, 14). These results indicate that although 17q is involved frequently in copy number alterations in several cancers, the target loci and genes might be different from one tumor type to another.

In summary, the present study demonstrates that although the 17q region contains hundreds of genes, only three genes were frequently amplified and overexpressed in gastric cancers, as compared with normal gastric epithelial tissues. The consistent overexpression of *TOP2A* in our gastric cancers suggests that this gene may be a potential target for topo2 inhibitors in gastric cancer patients. The overexpression of EST AA552509, in the majority of our samples, suggests that this novel gene may play a critical role in gastric tumorigenesis. We have initiated additional studies to explore the possible biological and clinical significance of these genes in gastric cancer development and progression.
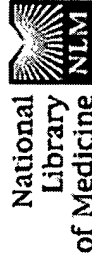
ACKNOWLEDGMENTS

REFERENCES

1. Whelan, S., Parkin, D., and Masuyer, E. Trends in Cancer Incidence and Mortality. IARC Scientific Publ. No. 102. Lyon, France: IARC, 1993.
2. Locke, G. R. R., Talley, N. J., Carpenter, H. A., Harmsen, W. S., Zinsmeister, A. R., and Melton, L. J. R. Changes in the site- and histology-specific incidence of gastric cancer during a 50-year period. Gastroenterology, *109:* 1750–1756, 1995.
3. Werner, M., Becker, K. F., Keller, G., and Hofler, H. Gastric adenocarcinoma: pathomorphology and molecular pathology. J. Cancer Res. Clin Oncol., *127:* 207–216, 2001.

4. Lin, W., Kao, H. W., Robinson, D., Kung, H. J., Wu, C. W., and Chen, H. C. Tyrosine kinases and gastric cancer. Oncogene, *19:* 5680–5689, 2000.

5. Allgayer, H., Babic, R., Gruetzner, K. U., Tarabichi, A., Schildberg, F. W., and Heiss, M. M. c-erbB-2 is of independent prognostic relevance in gastric cancer and is associated with the expression of tumor-associated protease systems. J. Clin. Oncol., *18:* 2201–2209, 2000.

6. El-Rifai, W., Frierson, H. J., Moskaluk, C., Harper, J., Petroni, G., Bissonette, E., Knuutila, S., and Powell, S. Genetic differences between adenocarcinomas arising in Barrett's esophagus and gastric mucosa. Gastroenterology, *121:* 592–598, 2001.

7. El-Rifai, W., Frierson, H. J., Harper, J., Powell, S., and Knuutila, S. Expression profiling of gastric adenocarcinoma using cDNA array. Int. J. Cancer, *92:* 828–832, 2001.

8. Monni, O., Barlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. Proc. Natl. Acad. Sci. USA, *98:* 5711–5716, 2001.

9. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat. Genet., *14:* 457–460, 1996.

10. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat. Genet., *23:* 41–46, 1999.

11. Mousses, S., Bittner, M., Chen, Y., Dougherty, E., Baxevanis, A., Meltzer, P., and Trent, J. (eds.). Functional Genomics, pp. 113–137. Oxford: Oxford University Press, 2000.

12. Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S. B., Pohida, T., Smith, P. D., Jiang, Y., Gooden, G. C., Trent, J. M., and Meltzer, P. S. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res., *58:* 5009–5013, 1998.

13. Chen, Y., Dougherty, E., and Bittner, M. Ratio-based decision and the quantitative analyses of cDNA microarray images. J. Biomed. Optics, *2:* 364–374, 1997.

14. Kauraniemi, P., Bärlund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. Cancer Res., *61:* 8235–8240, 2001.

15. Harris, L. N., Yang, L., Liotcheva, V., Pauli, S., Iglehart, J. D., Colvin, O. M., and Hsieh, T. S. Induction of topoisomerase II activity after ErbB2 activation is associated with a differential response to breast cancer chemotherapy. Clin. Cancer Res., *7:* 1497–1504, 2001.

16. Cuello, M., Ettenberg, S. A., Clark, A. S., Keane, M. M., Posner, R. H., Nau, M. M., Dennis, P. A., and Lipkowitz, S. Down-regulation of the erbB-2 receptor by trastuzumab (herceptin) enhances tumor necrosis factor-related apoptosis-inducing ligand-mediated apoptosis in breast and ovarian cancer cell lines that overexpress erbB-2. Cancer Res., *61:* 4892–4900, 2001.

17. Osheroff, N. DNA topoisomerases. Biochim. Biophys. Acta, *1400:* 1–2, 1998.

18. Piccart, M., Lohrisch, C., Di Leo, A., and Larsimont, D. The predictive value of HER2 in breast cancer. Oncology, *61* (Suppl. S2): 73–82, 2001.

19. Menard, S., Fortis, S., Castiglioni, F., Agresti, R., and Balsari, A. HER2 as a prognostic factor in breast cancer. Oncology, *61* (Suppl. S2): 67–72, 2001.

20. Jarvinen, T. A., Kononen, J., Pelto-Huikko, M., and Isola, J. Expression of topoisomerase IIα is associated with rapid cell proliferation, aneuploidy, and c-erbB2 overexpression in breast cancer. Am. J. Pathol., *148:* 2073–2082, 1996.

21. Smith, K., Houlbrook, S., Greenall, M., Carmichael, J., and Harris, A. L. Topoisomerase II α co-amplification with erbB2 in human primary breast cancer and breast cancer cell lines: relationship to m-AMSA and mitoxantrone sensitivity. Oncogene, *8:* 933–938, 1993.

22. Ross, J. S., and McKenna, B. J. The HER-2/neu oncogene in tumors of the gastrointestinal tract. Cancer Investig., *19:* 554–568, 2001.

23. Gurel, S., Dolar, E., Yerci, O., Samli, B., Ozturk, H., Nak, S. G., Gulten, M., and Memik, F. The relationship between c-erbB-2 oncogene expression and clinicopathological factors in gastric cancer. J. Int. Med. Res., *27:* 74–78, 1999.

24. Dankort, D., Maslikowski, B., Warner, N., Kanno, N., Kim, H., Wang, Z., Moran, M. F., Oshima, R. G., Cardiff, R. D., and Muller, W. J. Grb2 and Shc adapter proteins play distinct roles in Neu (ErbB-2)-induced mammary tumorigenesis: implications for human breast cancer. Mol. Cell. Biol., *21:* 1540–1551, 2001.

25. Shen, K., and Novak, R. F. DDT stimulates c-erbB2, c-met, and STATS tyrosine phosphorylation, Grb2-Sos association, MAPK phosphorylation, and proliferation of human breast epithelial cells. Biochem. Biophys. Res. Commun., *231:* 17–21, 1997.

26. Tari, A. M., and Lopez-Berestein, G. GRB2: a pivotal protein in signal transduction. Semin. Oncol., *28:* 142–147, 2001.

27. Houen, G. Mammalian Cu-containing amine oxidases (CAOs): new methods of analysis, structural relationships, and possible functions. APMIS Suppl., *96:* 1–46, 1999.

28. Hirst, J., Bright, N. A., Rous, B., and Robinson, M. S. Characterization of a fourth adapter-related protein complex. Mol. Biol. Cell, *10:* 2787–2802, 1999.

NCBI

PubMed

| Entrez | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals | Books |

National Library of Medicine NLM

Search PubMed

☑ Limits   for   Go Clear

Preview/Index   History   Clipboard   Details

About Entrez

Display Abstract   Show: 20   Sort   Send to Text

Text Version

Field: **Author**, Limits: **Publication Date from 2000 to 2000**

Entrez PubMed
Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities

☐ 1: Breast Cancer Res Treat. 2000 Sep;63(2):163-9.   Related Articles, Links

[SpringerLink]
FULL-TEXT ARTICLE

Relevance of p185 HER-2/neu oncoprotein quantification in human primary breast carcinoma.

Bermont L, Algros MP, Baron MH, Adessi GL.

PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
LinkOut
Cubby

Service d'Oncologie et d'Endocrinologie Moleculaires (C.H.U. Besancon), France.

Related Resources
Order Documents
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

The c-erbB-2 proto-oncogene encodes a transmembrane protein tyrosine kinase receptor of 185 kDa (p185) and has been associated with several types of human cancers. In human breast cancer, overexpression of p185 occurs in 15-30% of cases, correlates with poor prognostic factors and characterizes breast cancers with a more aggressive behavior. Overexpression of p185 is usually associated with c-erbB-2 amplification, though it may occur independently and thus define subpopulations of breast cancers which might be of clinical interest. p185 expression is usually detected by immunohistochemistry (IHC) and few studies have been carried out to evaluate the p185 content of breast cancers with an ELISA technique. In this context, we showed, in 106 breast cancer samples, that p185 was expressed at high levels in 13.2%, intermediate levels in 55.7% and negative ones in 31.1% of cases. All p185 positive samples showed a c-erbB-2 oncogene amplification while none of the p185 negative samples and only 4% of p185 intermediate samples had an amplification of c-erbB-2. p185 expression is significantly correlated with the negativity of estrogen and progestrone receptors, with high levels of cathepsin D and in some conditions with axillary nodal involvement. Thus, using the p185 ELISA assay, the c-erbB-2 status of breast cancers can be defined and moreover a subset can be discriminated which is characterized by intermediate levels of p185 and absence of c-erbB-2 amplification. The quantitative approach towards p185 in breast cancers affords the possibility of identifying more appropriately patients with high or low risk and thus permits adaptation of therapeutic regimens.

PMID: 11097092 [PubMed - indexed for MEDLINE]

Display | Abstract

Show: 20 | Sort | Send to | Text

Oct 29 2004 13:56:18

Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer

Profiling of Differentially Expressed Cancer-related Genes in Esophageal Squamous Cell Carcinoma (ESCC) Using Human Cancer cDNA Arrays: Overexpression of Oncogene *MET* Correlates with Tumor Differentiation in ESCC

Ying Chuan Hu,[1] King Yin Lam, Simon Law, John Wong, and Gopesh Srivastava[2]

Departments of Pathology [Y. C. H., K. Y. L., G. S.] and Surgery [S. L., J. W.], Faculty of Medicine, The University of Hong Kong, Hong Kong

ABSTRACT

Purpose: To examine the global gene expression of cancer-related genes in esophageal squamous cell carcinoma (ESCC) through the use of Atlas Human Cancer Array membranes printed with 588 well-characterized human genes involved in cancer and tumor biology.

Experimental Design: Two human ESCC cell lines (HKESC-1 and HKESC-2) and one morphologically normal esophageal epithelium tissue specimen from the patient of which the HKESC-2 was derived were screened in parallel using cDNA expression arrays. The array results were additionally validated using semiquantitative PCR. The overexpression of oncogene *MET* was studied more extensively for its protein expression by immunohistochemistry in the two ESCC cell lines and their corresponding primary tissues and 61 primary ESCC resected specimens. Sixteen of these 61 ESCC cases also had available the corresponding morphologically normal esophageal epithelium tissues and were also analyzed for MET expression. The clinicopathological features associated with overexpression of the *MET* gene were also correlated.

Results: The results of cDNA arrays showed that 13 cancer-related genes were up-regulated ≥2-fold (CDC25B, cyclin D1, PCNA, MET, Jagged 2, Integrin α3, Integrin α6, Integrin β4, Caveolin-2, Caveolin-1, MMP13, MMP14, and BIGH3) and 5 genes were down-regulated ≥2-fold (CK4, Bad, IGFBP2, CSPCP, and IL-1RA) in both ESCC cell lines at the mRNA level. Semiquantitative RT-PCR analysis of 9 of these differentially expressed genes, including the *MET*

gene, gave results consistent with cDNA array findings. The immunostaining results of the expression of *MET* gene showed that MET was overexpressed in both ESCC cell lines and their corresponding primary tumors at the protein level, validating the cDNA arrays findings. The results of the clinical specimens showed that the *MET* gene was overexpressed in ESCC compared with normal esophageal epithelium in 56 of 61 cases (92%). Moreover, the overexpression of MET protein was more often seen in well/moderately differentiated than in poorly differentiated ESCC.

Conclusions: Multiple cancer-related genes are differentially expressed in ESCC, the oncogene *MET* is overexpressed in ESCC compared with normal esophageal epithelium, and its protein overexpression correlates with tumor differentiation in ESCC.

INTRODUCTION

Despite advances in multimodality therapy, the prognosis for patients with ESCC[3] still remains poor, with an average 5-year survival rate <10% (1–5). The development of new treatment modalities, diagnostic technologies, and preventive approaches will require a better understanding of the molecular mechanisms underlying esophageal carcinogenesis. We have demonstrated earlier that cDNA arrays technology is a very useful tool for identifying differentially expressed genes in ESCC and reported the detection of 61 differentially expressed genes of 588 genes studied using Atlas Human cDNA Expression Arrays (6). In the present study, we specifically examined the global gene expression of cancer-related genes involved in the pathogenesis of ESCC by using the Human Cancer Array membranes printed with 588 well-characterized human genes involved in cancer and tumor biology. Among them, 235 genes were the same as those on Atlas Human cDNA Expression Arrays (6). The cancer-related genes analyzed in the present study are divided into six groups: (*a*) cell cycle regulators, growth regulators, and intermediate filament markers; (*b*) apoptosis, oncogenes, and tumor suppressors; (*c*) DNA damage response/repair and recombination; cell fate and development; and receptors; (*d*) cell adhesion and motility; and angiogenesis; (*e*) invasion regulators and cell-cell interactions; and (*f*) growth factors and cytokines. Using the Atlas Human cDNA Expression Arrays, 18 of 588 cancer-related genes examined were

---

---

[3] The abbreviations used are: ESCC, esophageal squamous cell carcinoma; CK, cytokeratin; CSPCP, cartilage-specific proteoglycan core protein; IHC, immunohistochemistry; IL-1RA, interleukin 1 receptor antagonist; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; MMP, matrix metalloproteinase; RT-PCR, reverse transcription-PCR.

identified to be differentially expressed in ESCC. Importantly, the mRNA of oncogene *MET* was found to be overexpressed in both the newly established ESCC cell lines HKESC-1 and HKESC-2. This prompted us to additionally examine its protein expression in the cell lines, their respective primary tissues, a large series of primary ESCC tumors, and corresponding morphologically normal esophageal epithelium tissues. Furthermore, we also analyzed the relationships between *MET* expression and the clinicopathological parameters of ESCC.

MATERIALS AND METHODS

ESCC Cell Lines and Control Specimen. Two human ESCC cell lines (HKESC-1 and HKESC-2) and one morphologically normal esophageal epithelium tissue specimen from the patient of which the HKESC-2 was derived were used for the Human Cancer cDNA Expression Arrays experiment (6). Both cell lines were established from Hong Kong Chinese patients with moderately differentiated ESCC: HKESC-1 from a 47-year-old man and HKESC-2 from a 46-year-old woman (7, 8). Both cell lines grew as adherent monolayers and cultured in Minimum Essential Medium with non-essential amino acids (MN) (Sigma, Saint Louis, MO) medium containing 10% fetal bovine serum (7, 8). Cells were harvested from passage 31 of HKESC-1 and passage 4 of HKESC-2 at 80–90% confluency, respectively. Unfortunately, the collected normal esophageal epithelium tissue from the patient of which the HKESC-1 was derived could not be used as a control, because the specimen was too small, and only a small amount of RNA could be extracted from it.

Human Cancer cDNA Arrays, Probes, Hybridization, and Data Analysis. Atlas Human Cancer cDNA Expression Arrays membranes used in this study were purchased from Clontech (Palo Alto, CA). The membrane contained 10 ng of each gene-specific cDNA from 588 known cancer-related genes and 9 housekeeping genes (Fig. 1). Several plasmid and bacteriophage DNAs and blank spots are also included as negative and blank controls to confirm hybridization specificity. A complete list of the 588 cancer-related genes with array positions and GenBank accession number of the Atlas Human Cancer Expression Arrays used here can be accessed at the website.[4]

Total RNA was extracted using the TRIzol reagent protocol (Life Technologies, Inc., Gaithersburg, MD) from the two ESCC cell lines (HKESC-1 and HKESC-2) and one corresponding morphologically normal esophageal epithelium from the patient of which the HKESC-2 was derived. mRNA was then isolated from the total RNA using the Straight A's mRNA Isolation System (Novagen, Madison, WI). The ^{32}P-labeled cDNA probes were generated by reverse transcription of 1 µg of mRNA of each sample in the presence of [α-^{32}P]dATP. Equal amounts of cDNA probes (3 × 10^6 cpm/µl) from the ESCC cell lines and normal esophageal epithelium were then hybridized to separate Atlas Human Cancer cDNA array membranes for 24 h at 42°C and washed according to the supplier's instructions. The array membranes were then exposed to X-ray film at −70°C for 2–5 days. Autoradiographic intensity was analyzed using AtlasImage analysis software
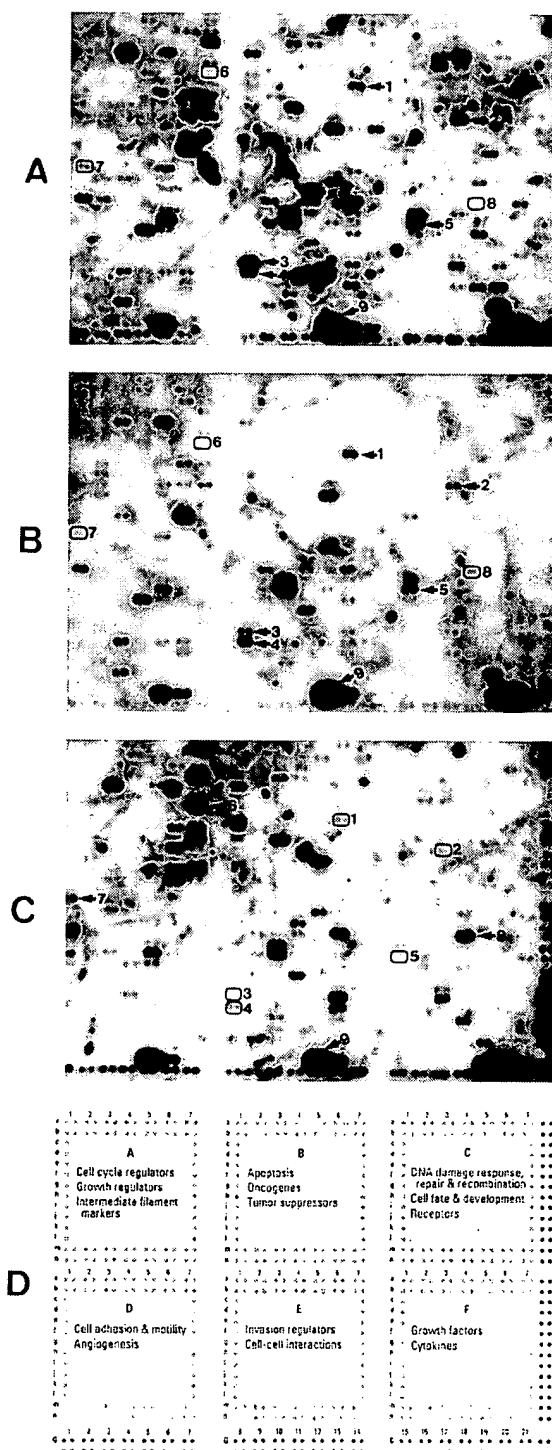
[4] Internet address: http://www.clontech.com/atlas/genelists/Hcancer.xls.

Fig. 1 *A–C*, global gene expression profiles of cancer-related genes of two human ESCC cell lines HKESC-1 (*A*) and HKESC-2 (*B*) and one corresponding morphologically normal esophageal epithelium (*C*) from the patient of which the HKESC-2 was derived using Atlas Human Cancer cDNA Expression Arrays. Some of the differentially expressed genes are indicated: *1*, MET (B6h); *2*, Jagged 2 (C3k); *3*, MMP13 (E1j); *4*, MMP14 (E1k); *5*, BIGH3 (F1f); *6*, CK4 (A7g); *7*, CSPCP (D1a); *8*, IL-1RA (F4d); and *9*, GAPDH (G12). *D*, schematic diagram of Atlas

Table 1 Summary of differentially expressed genes in both ESCC cell lines HKESC-1 and HKESC-2 when compared with one corresponding morphologically normal esophageal epithelium tissue specimen (N) from the patient of which the HKESC-2 was derived by Atlas Human Cancer cDNA Expression Arrays

Arrays			Intensity ratio	
Position	Name of gene	Chromosome location	HKESC-1/N	HKESC-2/N
A1k	CDC25B	20p13	2.0	2.1
A2l	Cyclin D1	11q13	2.2	2.0
A5e	PCNA	20pter-p12	2.2	2.6
B6h	MET	12p12.1	2.8	2.9
C3k	Jagged 2	14q32	3.7	3.4
D3k	Integrin α3	17	2.6	2.5
D3n	Integrin α6	2	3.3	3.2
D4g	Integrin β4	17q11-qter	2.4	2.7
D6k	Caveolin-2	7q31.1-q31.2	8.8	5.2
D6l	Caveolin-1	7q31.1	9.5	7.2
E1j	MMP13	11q22.3	32.3	29.8
E1k	MMP14	14q11-q12	4.5	4.9
F1f	BIGH3	5q31	2.2	2.3
A7g	CK4	12q13	1/5.0	1/6.9
B1i	Bad	11	1/2.1	1/4.3
C6c	IGFBP2	2q33-q34	1/8.9	1/2.5
D1a	CSPCP	15q26.1	1/3.0	1/2.2
F4d	IL-1RA	2q14.2	1/4.4	1/2.2

(version 1.01; Clontech). The signal intensities were normalized by comparing the expression of housekeeping genes Ubiquitin (G5) and GAPDH (G12) and calculated as described previously (6). Housekeeping genes Ubiquitin and GAPDH were selected for normalization, because their expression was constant in this Cancer Array hybridization system. Genes were considered to be up-regulated when the intensity ratio between expression in the ESCC cell lines compared with normal esophageal epithelium was ≥2-fold. Genes were labeled as down-regulated when the ratio between normal and ESCC cell lines was ≥2-fold. To test the reproducibility of Cancer Array hybridization system, we repeated hybridization using new probes generated from the original mRNA, which gave similar results.

Semiquantitative RT-PCR. cDNA was generated using 1 μg of total RNA from the two ESCC cell lines (HKESC-1 and HKESC-2) and one corresponding morphologically normal esophageal epithelium from the patient of which the HKESC-2 was derived as template and 2.5 mM oligo d(T)$_{16}$ primers in a 20-μl reaction mixture, and the reverse transcription was carried out at 42°C for 1 h followed by 95°C for 10 min using the GeneAmp RNA PCR Core kit (Perkin-Elmer, Branchburg, NJ).

cDNA (2 μl) was amplified in a 25-μl PCR reaction mixture containing 1× PCR buffer, (10 mM Tris-HCl, pH 8.3, 50 mM KCl) 1.9 or 2.4 mM of MgCl$_2$, 0.5 μM of primers, 0.18 mM of deoxynucleotide triphosphate, and 1 unit AmpliTaq Gold DNA Polymerase. The hot-start PCR reaction was as follows: 95°C for 10 min followed by 25–40 cycles of 1 min denaturation at 94°C, 1 min annealing at 60°C (for primers of MET, Jagged, MMP13, MMP14, BIGH3, CK1, CK4, IL-1RA, and GAPDH) or 50°C (for primers of CSPCP), and 1 min extension at 72°C. The final step of extension was for 10 min at 72°C. The PCR reagents were purchased from Perkin-Elmer.

The sequences of gene specific primers for RT-PCR were the same as those of Cancer cDNA arrays (data not shown because of the copyright agreement by Clontech, Palo Alto, CA) except for the primers specific for MET, which were the same as described before (9). All of the primers were synthesized by Integrated DNA Technologies Inc., Coralville, IA. The cycle number was optimized for each gene-specific primer pair to ensure that amplification was in the linear range, and the results were semiquantitative. PCR product (12 μl) was visualized by electrophoresis on a 2% agarose gel stained with ethidium bromide and quantitated by densitometry using a dual-intensity transilluminator equipped with Gelworks 1D Intermediate software (version 2.51).

Collection of Tissues and Clinicopathological Data. The tissues were obtained from 61 (50 men and 11 women) patients with ESCC resected between 1996 and 1998 in Queen Mary Hospital, The University of Hong Kong. The patient ages ranged from 41 to 83 years, with a mean age of 65 years. The specimens were dissected and examined in the fresh state. Representative tissue specimens from tumors and matching normal esophageal epithelium tissues were snap-frozen in liquid nitrogen and stored at −80°C. Other representative blocks were taken and processed in paraffin for histological examination. The carcinomas were found in the upper (n = 10, 16%), middle (n = 35, 57%), and lower (n = 16, 26%) third of the esophagus. The median length of the tumors was 5.5 cm (range, 1–11). The histology of the carcinomas was reviewed according to the criteria described previously (4). The ESCC tumors were well differentiated in 20 (33%) cases, moderately differentiated in 29 (48%), and poorly differentiated in 12 (20%). The carcinomas were staged according to the Tumor-Node-Metastasis classification (10). Many tumors were stage III (n = 35, 57%) or II (n = 23, 38%); of the remainder, 1 was stage I, and 2 were stage IV.

IHC Staining of MET Gene. Expression of the MET gene was investigated by streptavidin-biotin-peroxidase complex method. Briefly, 6-μm frozen sections were cut from two pellets harvested from cultured cell lines HKESC-1 and HKESC-2, the cell lines corresponding primary tissues and 61 primary ESCC tumors. Sixteen of these 61 ESCC cases also had available the corresponding morphologically normal esophageal epithelium tissues and were also analyzed for MET expression. After endogenous peroxidase activity was quenched and nonspecific binding was blocked, polyclonal rabbit anti-MET antibody (Santa Cruz Biotechnology, Santa Cruz, CA) was incubated at 4°C overnight at a dilution of 1:50. The secondary antibody was biotinylated swine antirabbit antibody (DAKO, Glostrup, Denmark) used at a dilution of 1:200 for 30 min at 37°C. After washing, sections were incubated with StreptAB-Complex/horseradish peroxidase (DAKO; 1:100 dilution) for 30

A

PCR cycle numbers

20 25 30 35 40 (−)

MET ← 656bp

Jagged 2 ← 219bp

MMP13 ← 298bp

MMP14 ← 387bp

BIGH3 ← 244bp

CK1 ← 230bp

CK4 ← 290bp

CSPSP ← 251bp

IL-1RA ← 295bp

GAPDH ← 269bp

B

HKESC-1 HKESC-2 Normal (−)

MET ← 656bp

Jagged 2 ← 219bp

MMP13 ← 298bp

MMP14 ← 387bp

BIGH3 ← 244bp

CKI ← 230bp

CK4 ← 290bp

CSPCP ← 251bp

IL-1RA ← 295bp

GAPDH ← 269bp

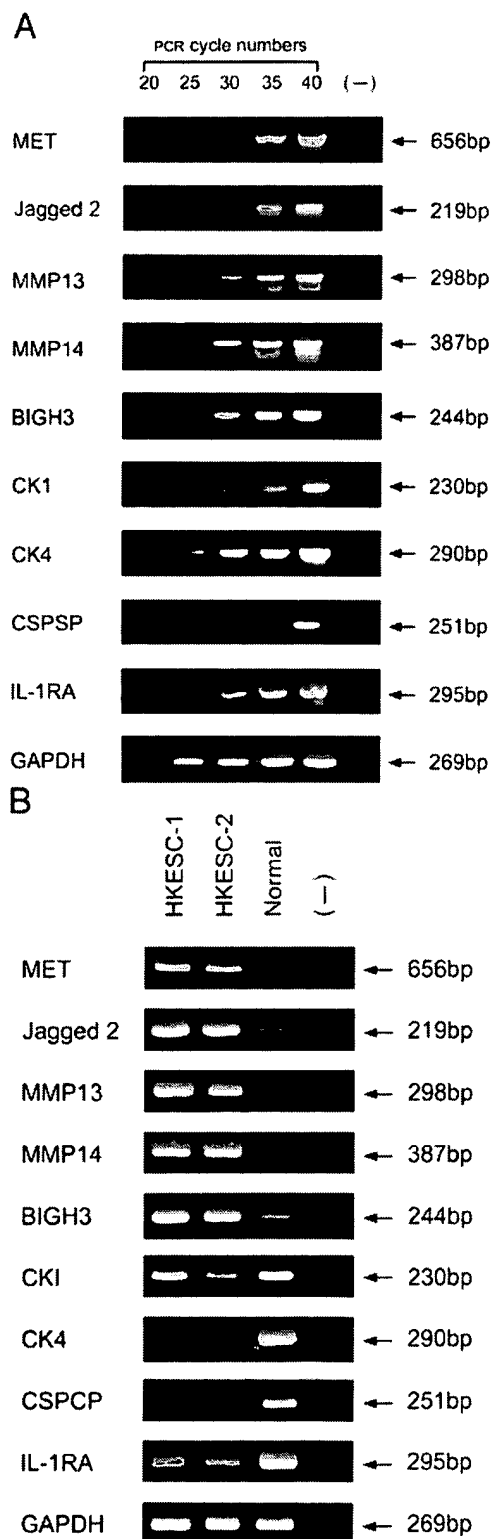*Fig. 2* RT-PCR analysis of *MET, Jagged 2, MMP13, MMP14, BIGH3, CKI, CK4, CSPCP, IL-1RA,* and *GAPDH* genes in ESCC cell lines HKESC-1 and HKESC-2 and one corresponding morphologically normal esophageal epithelium (*Normal*) from the patient of which

min at 37°C. Negative controls were performed by replacing the primary antibody by normal serum. Each section was independently assessed by two histopathologists (Y. C. H. and K. Y. L.) without previous knowledge of the other data of the patients. All of the fields in the selected block were taken into consideration for assessment of immunostaining. The percentage of tumor cells stained of total tumor cells noted was reported. Representative areas of each section were selected, and cells were counted in at least four fields (at ×200). Scoring was based on the percentage of positive cells. The IHC staining was identified as (−): no expression; (+): <10% of cells were stained; (2+): 10–50% of cells stained; (3+): >50% of cells stained; (2+) − (3+) was defined as overexpression.

Statistical Analysis. Comparisons between groups were performed using the χ^2 test and t test when appropriate. $P < 0.05$ was used to determine statistical significance. All of the statistical tests were performed with the GraphPad Prism software version 3.0 (GraphPad Software, Inc., San Diego, CA).

RESULTS

Identification of Differentially Expressed Cancer-related Genes in ESCC by the Human Cancer cDNA Arrays. The general expression profiles of 588 cancer-related genes in HKESC-1 (Fig. 1*A*), HKESC-2 (Fig. 1*B*), and normal esophageal epithelium tissue (Fig. 1*C*) as determined by the Human Cancer cDNA Arrays are shown in Fig. 1. No signals were visible in the blank spots (G1, G8, and G15) and negative control spots (G2–4, G9–11, and G16–18; Fig. 1) indicating that the Cancer Arrays hybridization was highly specific. The comparison of the autoradiographic intensities between ESCC cell lines and normal esophageal epithelium showed that 13 genes were up-regulated and 5 genes down-regulated ≥2-fold in both cell lines (Table 1).

Confirmation of Differentially Expressed Cancer-related Genes by Semiquantitative RT-PCR. The semiquantitative RT-PCR results showed that *MET, Jagged 2, MMP13, MMP14,* and *BIGH3* genes were up-regulated in cell lines HKESC-1 and HKESC-2, whereas *CK4, CSPCP,* and *IL-1RA* were down-regulated in HKESC-1 and HKESC-2 (Fig. 2). *CK1* was down-regulated in HKESC-2 but not in HKESC-1. These results are similar to those detected by Human Cancer cDNA Arrays (Fig. 1).

Expression of *MET* Gene in ESCC Cell Lines and Their Respective Primary Tissues. The immunostaining results of MET expression in ESCC cell lines HKESC-1 and HKESC-2

HKESC-2 was derived. *A*, determination of optimal number of PCR cycles for different gene-specific primer pairs. mRNA from HKESC-1 was used to determine the optimal number of PCR cycles for genes *MET, Jagged 2, MMP13, MMP14, BIGH3,* and *GAPDH.* mRNA from the normal esophageal epithelium was used to determine optimal number of PCR cycles for genes *CK1, CK4, CSPCP,* and *IL-1RA. B,* expression of *MET* (31 cycles), *Jagged 2* (31 cycles), *MMP13* (31 cycles), *MMP14* (31 cycles), *BIGH3* (31 cycles), *CK1* (33 cycles), *CK4* (28 cycles), *CSPCP* (40 cycles), *IL-1RA* (28 cycles), and *GAPDH* (25 cycles) genes in two ESCC cell lines HKESC-1 and HKESC-2 and one corresponding morphologically normal esophageal epithelium (*Normal*) from the patient of which the HKESC-2 was derived. PCR products were electrophoresed on 2% agarose gel containing ethidium bromide.

Table 2 Results of immunostaining in ESCC cell lines and their
respective primary tissues

	Cell lines[a]		Primary tissues			
	HKESC-1	HKESC-2	T1[b]	N1	T2	N2
MET	+++	+++	+++	+	+++	+

[a] Cell lines HKESC-1 and HKESC-2 were established from T1 and
T2 tissues, respectively.

[b] T, ESCC; N, morphologically normal esophageal epithelium.

Table 3 Summary of IHC staining results in clinical ESCC tumors
and morphologically normal esophageal epithelium tissues

	MET expression				
Diagnosis	−	+	++	+++	P
Normal (n = 16)	1	10	5	0	<0.0001
Carcinoma (n = 61)	5	0	11	45	
Well (n = 20)	0	0	4	16	
Moderate (n = 29)	0	0	4	25	<0.0001
Poor (n = 12)	5	0	3	4	

and their corresponding primary tissues were summarized in
Table 2. MET protein was found to be overexpressed in both the
ESCC cell lines (HKESC-1 and HKESC-2) and the primary
tumors from which these cell lines were established (Table 2).

**Expression of *MET* Gene in Primary ESCC Tumors
and Morphologically Normal Esophageal Epithelium Tis-
sues.** The IHC staining results of MET expression in 61 pri-
mary ESCC tumors and 16 corresponding morphologically nor-
mal esophageal epithelium tissues are summarized in Table 3
and shown in Fig. 3. As shown in Table 3, MET was overex-
pressed in ESCC in 56 of 61 cases (92%). MET protein had a
significantly higher incidence of overexpression in ESCCs than
morphologically normal esophageal epithelium tissues (P <
0.0001; Table 3). The expression of MET protein was localized
in the cytoplasm and cell membrane of tumor cells (Fig. 3).

**The Clinicopathological Features Associated with
Overexpression of the *MET* Gene.** The clinicopathological
features of cases showing MET overexpression and negative
cases of primary ESCC were compared in Table 4. MET over-
expression had significant correlation with ESCC differentiation
(P < 0.0001) but had no relationship with the patient gender,
age, tumor size, site, or stage (Table 4). The well/moderately
differentiated ESCC showed more intense expression of MET
than poorly differentiated ones (P < 0.0001; Table 3).

DISCUSSION

We have demonstrated previously that the cDNA arrays is
a very powerful tool for identifying differentially expressed
genes in ESCC (6), because this approach permits the investi-
gation of hundreds of genes simultaneously in one experiment.
In the current study, we have used the Human Cancer cDNA
Expression Arrays to specifically study the global differential
expression of cancer-related genes in two human ESCC cell
lines (HKESC-1 and HKESC-2). We have identified 18 cancer-
related genes differentially expressed in both of these ESCC cell
lines, 13 of which were up-regulated ≥2-fold (CDC25B, cyclin
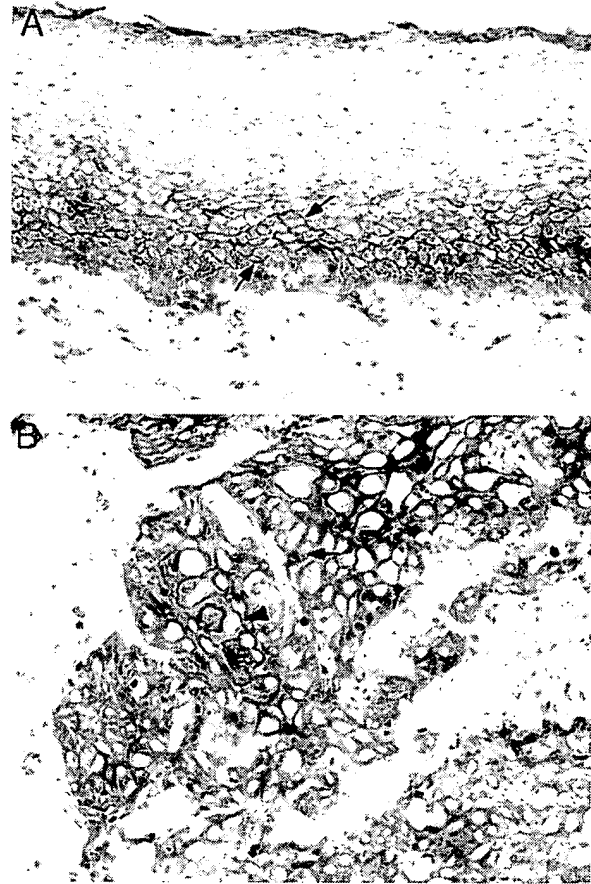D1, PCNA, MET, Jagged 2, Integrin α3, Integrin α6, Integrin

Fig. 3 Photomicrographs of MET expression by IHC staining in mor-
phologically normal esophageal epithelium and ESCC. *A*, MET IHC in
morphologically normal esophageal epithelium showing MET expres-
sion was restricted to the parabasal cell layer (*arrow*), 3,3'-diaminoben-
zidine; ×400. *B*, MET IHC in ESCC showing the cytoplasm (*arrow*)
and cell membrane (*arrowhead*) of most tumor cells are strongly posi-
tive for MET, 3,3'-diaminobenzidine; ×500.

β4, Caveolin-2, Caveolin-1, MMP13, MMP14, and BIGH3) and
5 of which were down-regulated ≥2-fold (CK4, Bad, IGFBP2,
CSPCP, and IL-1RA) in both ESCC cell lines at mRNA level.
These results of the up-regulation of *CDC25B* and *cyclin D1*
genes in ESCC obtained in this study confirmed our earlier
results on the overexpression of these genes in ESCC, which
were obtained using Atlas Human cDNA expression arrays (6).
Subsequent RT-PCR analysis of 9 of these differentially ex-
pressed genes including *MET, Jagged 2, MMP-13, MMP-14,
BIGH3, CK4, CSPCP,* and *IL-1RA* confirmed the differential
profiles uncovered by Human Cancer cDNA arrays hybridi-
zation.

Some of these differentially expressed genes identified
here have been reported previously to be implicated in the
pathogenesis of other malignancies or esophageal cancer. For
example, it is well known that cyclin D1 is a key cell cycle
regulator in the G_1 to S phase progression; through a complex
with CDK4, it phosphorylates and inactivates retinoblastoma

Table 4 Clinicopathological features of MET overexpression and negative cases of primary ESCC

Clinicopathological features	MET overexpression		P
	+	−	
Male:female ratio	46:10	4:1	$P = 1.0000$ (Fisher's exact test)
Median age (yr.)	65.5	62.8	$P = 0.5540$ (t test)
Median tumor length (cm)	5.4	6.8	$P = 0.1209$ (t test)
Location			
Upper	10	0	$P = 0.4821$ (χ^2 test)
Middle	31	4	
Lower	15	1	
Differentiation			
Well	20	0	$P < 0.0001$ (χ^2 test)
Moderate	29	0	
Poor	7	5	
Stage			
I/II	22	2	$P = 1.0000$ (Fisher's exact test)
III/IV	34	3	

gene protein. The abnormalities of proto-oncogene cyclin D1 have been implicated in the tumorigenesis of numerous tumor types including ESCC (11). Previously, overexpression and/or amplification of cyclin D1 has been consistently found in ESCC (11–13). In this study, the mRNA of cyclin D1 showed overexpression in the two ESCC cell lines. These indicate that cyclin D1 overexpression is a very common molecular event in ESCC and may play an important role in the carcinogenesis of ESCC. In addition, our Cancer Array hybridization results demonstrated that several genes related to cell adhesion and invasion were overexpressed in HKESC-1 and HKESC-2. Although integrin α6 has been shown to be overexpressed in esophageal cancer (14), the expression of integrin α3, integrin β4, MMP13, or MMP14, to our knowledge, has not been reported before in ESCC. The integrins are major adhesion-receptor proteins that mediate cell migration and invasion. The MMP family has been shown to be involved in proteolytic degradation of the extracellular matrix to enhance tumor cell movement. The identification of these novel molecular alterations provided promising targets for assessment of invasion and metastatic potential of ESCC in the future.

MET oncogene was originally identified as a tumor-transforming gene (15, 16). It is located on chromosome 7q31 (15). This oncogene encodes a M_r 190,000 tyrosine kinase receptor for hepatocyte growth factor (17). A vast body of clinical and experimental data has demonstrated that the *MET* oncogene plays a crucial role in tumorigenesis of many tumors. *MET* gene has been found to be overexpressed in thyroid carcinomas (18, 19), gastric carcinomas and colorectal carcinomas (18, 19), ovarian carcinomas (20), endometrial carcinomas (21), pancreatic carcinomas (22, 23), renal cell carcinomas (24, 25), breast carcinomas (26–28), and prostatic carcinomas (29). These findings suggested that increased expression of the *MET* oncogene in human tumors might confer a selective growth advantage to tumor cells. However, information about MET expression in ESCC is very limited. An earlier study has indicated that MET mRNA was overexpressed in ESCC (30), but there has been no information about MET expression at the protein level in ESCC.

In this study, the Human Cancer cDNA arrays hybridization revealed that oncogene *MET* mRNA was expressed at a much higher level in ESCC than in normal tissue. Subsequent RT-PCR analysis additionally confirmed the findings from the Cancer cDNA arrays. With IHC, the majority of ESCC (56/61, 92%) was found to have significantly enhanced expression of MET compared with morphologically normal esophageal epithelium ($P < 0.0001$). Also, the findings provided additional evidence that MET mRNA was overexpressed during the development of ESCC.

In the current study, there was significant correlation between MET overexpression and ESCC differentiation ($P < 0.0001$). The well- or moderately differentiated ESCC had much more elevated MET expression than the poorly differentiated ones. This is in keeping with previous findings in other tumors (20, 31, 32). Di Renzo *et al.* (20) found MET to be most overexpressed in differentiated ovarian carcinomas. Huntsman *et al.* (31) observed that MET expression was enhanced in most benign ovarian tumors and appeared to be maximally overexpressed in borderline tumors and well-differentiated ovarian carcinomas. In renal cell carcinoma, a close relationship was observed between MET overexpression and the chromophilic subtype with a papillary growth pattern (32). However, in a number of tissues, MET becomes increasingly overexpressed as tumors become poorly differentiated (33). These combined findings suggest that the relationship of MET expression to tumor differentiation seems to vary among different tumor types.

In this study, MET protein was found to be overexpressed in both ESCC cell lines (HKESC-1 and HKESC-2) and the primary tumors from which these cell lines were established. This demonstrated that MET protein is overexpressed *in vitro* and *in vivo* in ESCC. More extensive examination in 61 cases of surgically resected ESCC samples provided additional evidence that the majority of ESCC tumors had MET overexpression in the natural history of ESCC development. The *MET* oncogene can be activated by overexpression (17), gene rearrangements (15), or mutations (34). Thus, the observed MET overexpression in ESCC in this study can be presumed to lead to MET activation and play a role in the pathogenesis of ESCC.

In conclusion, 18 cancer-related genes of 588, including MET, were identified to be differentially expressed in HKESC-1 and HKESC-2. Among these for the first time MET protein was

noted to be overexpressed in ESCC as compared with morphologically normal esophageal epithelium tissues, and the overexpression of MET was found to correlate with tumor differentiation in ESCC. These findings suggest that the activation of MET oncogene via overexpression might be important in the pathogenesis of ESCC.

REFERENCES

1. Blot, W. J. Esophageal cancer trends and risk factors. Semin. Oncol., 21: 403–410, 1994.

2. Lam, K. Y., and Ma, L. Pathology of esophageal cancers: local experience and current insights. Chin. Med. J. (Engl. Ed.), 10: 459–464, 1997.

3. Law, S. Y., Fok, M., Cheng, S. W., and Wong, J. A comparison of outcome after resection for squamous cell carcinomas and adenocarcinomas of the esophagus and cardia. Surg. Gynecol. Obstet., 175: 107–112, 1992.

4. Lam, K. Y., Tsao, S. W., Zhang, D., Law, S., He, D., Ma, L., and Wong, J. Prevalence and predictive value of p53 mutation in patients with esophageal squamous cell carcinomas: a prospective clinico-pathological study and survival analysis of 70 patients. Int. J. Cancer, 74: 212–219, 1997.

5. Montesano, R., Hollstein, M., and Hainaut, P. Genetic alterations in esophageal cancer and their relevance to etiology and pathogenesis: a review. Int. J. Cancer, 69: 225–235, 1996.

6. Hu, Y. C., Lam, K. Y., Law, S., Wong, J., and Srivastava, G. Identification of differentially expressed genes in esophageal squamous cell carcinoma (ESCC) by cDNA expression arrays: overexpression of Fra-1, Neogenin, Id-1, and CDC25B genes in ESCC. Clin. Cancer Res., 7: 2213–2221, 2001.

7. Hu, Y. C., Lam, K. Y., Wan, T. S. K., Fang, W. G., Ma, E. S. K., Chan, L. C., and Srivastava, G. Establishment and characterization of HKESC-1, a new cancer cell line from human esophageal squamous cell carcinoma. Cancer Genet. Cytogenet., 118: 112–120, 2000.

8. Hu, Y. C., Lam, K. Y., Law, S. Y., Wan, T. S. K., Ma, E. S. K., Kwong, Y. L., Chan, L. C., Wong, J., and Srivastava, G. Establishment, characterization, karyotyping, and CGH analysis of HKESC-2 and HKESC-3, two newly established human esophageal squamous cell carcinoma cell lines. Cancer Genet. Cytogenet., in press, 2001.

9. Klominek, J., Baskin, B., Liu, Z., and Hauzenberger, D. Hepatocyte growth factor/scatter factor stimulates chemotaxis and growth of malignant mesothelioma cells through c-met receptor. Int. J. Cancer, 76: 240–249, 1998.

10. Beahrs, O. H., Henson, D. E., Hutter, R. V. P., and Kennedy, B. J. (eds.). Manual for Staging of Cancer. 4th ed. pp. 57–59. Philadelphia, PA: J. B. Lippincott, 1992.

11. Sheyn, I., Noffsinger, A. E., Heffelfinger, S., Davis, B., Miller, M. A., and Fenoglio, P. C. Amplification and expression of the cyclin D1 gene in anal and esophageal squamous cell carcinomas. Hum. Pathol., 28: 270–276, 1997.

12. Jiang, W., Kahn, S. M., Tomita, N., Zhang, Y. J., Lu, S. H., and Weinstein, I. B. Amplification and expression of the human cyclin D1 gene in esophageal cancer. Cancer Res., 52: 2980–2983, 1992.

13. Jiang, W., Zhang, Y. J., Kahn, S. M., Hollstein, M. C., Santella, R. M., Lu, S. H., Harris, C. C., Montesano, R., and Weinstein, I. B. Altered expression of the cyclin D1 and retinoblastoma genes in human esophageal cancer. Proc. Natl. Acad. Sci. USA, 90: 9026–9030, 1993.

14. Tanaka, Y., Mimore, K., Shiraishi, T., Ohkura, Y., Takubo, K., Mafune, K., Barnard, G. F., and Mori, M. α6 integrin expression in esophageal carcinoma. Int. J. Oncol., 16: 725–729, 2000.

15. Cooper, C. S., Park, M., Blair, D. G., Tainsky, M. A., Huebner, K., Croce, C. M., and Vande, W. G. Molecular cloning of a new transforming gene from a chemically transformed human cell line. Nature (Lond.), 311: 29–33, 1984.

16. Park, M., Dean, M., Cooper, C. S., Schmidt, M., O'Brien, S. J., Blair, D. G., and Vande, W. G. Mechanism of met oncogene activation. Cell, 45: 895–904, 1986.

17. Giordano, S., Ponzetto, C., Di Renzo, M., Cooper, C. S., and Comoglio, P. M. Tyrosine kinase receptor indistinguishable from the c-met protein. Nature (Lond.), 339: 155–156, 1989.

18. Giordano, S., Di Renzo, M., Olivero, M., Mondino, A., Zhen, Z., Medico, E., and Comoglio, P. M. The c-met/HGF receptor in human tumours. Eur. J. Cancer Prev., 1 (Suppl 3): 45–49, 1992.

19. Prat, M., Narsimhan, R. P., Crepaldi, T., Nicotra, M. R., Natali, P. G., and Comoglio, P. M. The receptor encoded by the human c-MET oncogene is expressed in hepatocytes, epithelial cells and solid tumors. Int. J. Cancer, 49: 323–328, 1991.

20. Di Renzo, M., Olivero, M., Katsaros, D., Crepaldi, T., Gaglia, P., Zola, P., Sismondi, P., and Comoglio, P. M. Overexpression of the Met/HGF receptor in ovarian cancer. Int. J. Cancer, 58: 658–662, 1994.

21. Wagatsuma, S., Konno, R., Sato, S., and Yajima, A. Tumor angiogenesis, hepatocyte growth factor, and c-Met expression in endometrial carcinoma. Cancer (Phila.), 82: 520–530, 1998.

22. Furukawa, T., Duguid, W. P., Kobari, M., Matsuno, S., and Tsao, M. S. Hepatocyte growth factor and Met receptor expression in human pancreatic carcinogenesis. Am. J. Pathol., 147: 889–895, 1995.

23. Kiehne, K., Herzig, K. H., and Folsch, U. R. c-met expression in pancreatic cancer and effects of hepatocyte growth factor on pancreatic cancer cell growth. Pancreas, 15: 35–40, 1997.

24. Natali, P. G., Prat, M., Nicotra, M. R., Bigotti, A., Olivero, M., Comoglio, P. M., and Di Renzo, M. Overexpression of the met/HGF receptor in renal cell carcinomas. Int. J. Cancer, 69: 212–217, 1996.

25. Pisters, L. L., El, N. A., Luo, W., Malpica, A., and Lin, S. H. C-met proto-oncogene expression in benign and malignant human renal tissues. J. Urol., 158: 724–728, 1997.

26. Beviglia, L., Matsumoto, K., Lin, C. S., Ziober, B. L., and Kramer, R. H. Expression of the c-Met/HGF receptor in human breast carcinoma: correlation with tumor progression. Int. J. Cancer, 74: 301–309, 1997.

27. Byers, S., Park, M., Sommers, C., and Seslar, S. Breast carcinoma: a collective disorder. Breast Cancer Res. Treat., 31: 203–215, 1994.

28. Tuck, A. B., Park, M., Sterns, E. E., Boag, A., and Elliott, B. E. Coexpression of hepatocyte growth factor and receptor (Met) in human breast carcinoma. Am. J. Pathol., 148: 225–232, 1996.

29. Humphrey, P. A., Zhu, X., Zarnegar, R., Swanson, P. E., Ratliff, T. L., Vollmer, R. T., and Day, M. L. Hepatocyte growth factor and its receptor (c-MET) in prostatic carcinoma. Am. J. Pathol., 147: 386–396, 1995.

30. Porte, H., Triboulet, J. P., Kotelevets, L., Carrat, F., Prevot, S., Nordlinger, B., DiGioia, Y., Wurtz, A., Comoglio, P., Gespach, C., and Chastre, E. Overexpression of stromelysin-3, BM-40/SPARC, and MET genes in human esophageal carcinoma: implications for prognosis. Clin. Cancer Res., 4: 1375–1382, 1998.

31. Huntsman, D., Resau, J. H., Klineberg, E., and Auersperg, N. Comparison of c-met expression in ovarian epithelial tumors and normal epithelia of the female reproductive tract by quantitative laser scan microscopy. Am. J. Pathol., 155: 343–348, 1999.

32. Inoue, K., Karashima, T., Chikazawa, M., Iiyama, T., Yoshikawa, C., Furihata, M., Ohtsuki, Y., and Shuin, T. Overexpression of c-met proto-oncogene associated with chromophilic renal cell carcinoma with papillary growth. Virchows Arch., 433: 511–515, 1998.

33. Koochekpour, S., Jeffers, M., Rulong, S., Taylor, G., Klineberg, E., Hudson, E. A., Resau, J. H., and Vande, W. G. Met and hepatocyte growth factor/scatter factor expression in human gliomas. Cancer Res., 57: 5391–5398, 1997.

34. Jeffers, M., Schmidt, L., Nakaigawa, N., Webb, C. P., Weirich, G., Kishida, T., Zbar, B., and Vande, W. G. Activating mutations for the met tyrosine kinase receptor in human cancer. Proc. Natl. Acad. Sci. USA, 94: 11445–11450, 1997.